GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 17, 2003, 17:56:34 ; Search time 29.0718 Seconds
(without alignments)
12253.192 Million cell updates/sec

Title: US-09-914-286-1_COPY_5935_6180
Perfect score: 246
Sequence: 1 ctggcccatcagacgtcggc.....ccaaactcgccgtacacctg 246

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1029858 seqs, 724030393 residues

Total number of hits satisfying chosen parameters: 2059716

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
11: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
12: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
13: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
14: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	110.2	44.8	50937	9	US-09-808-880-1	Sequence 1, Appli
2	103.6	42.1	11220	9	US-09-860-846-32	Sequence 32, Appl
3	103.6	42.1	11220	9	US-09-988-384B-32	Sequence 32, Appl
4	103.6	42.1	11220	9	US-09-836-821-32	Sequence 32, Appl
5	103.6	42.1	11220	10	US-09-861-289-32	Sequence 32, Appl
6	103.6	42.1	36778	9	US-09-860-846-5	Sequence 5, Appli
7	103.6	42.1	36778	9	US-09-836-821-5	Sequence 5, Appli
8	103.6	42.1	36778	10	US-09-861-289-5	Sequence 5, Appli
9	103.6	42.1	37948	9	US-09-988-384B-5	Sequence 5, Appli
10	103.6	42.1	38506	9	US-09-793-708-19	Sequence 19, Appl
11	98.4	40.0	4689	9	US-09-860-846-34	Sequence 34, Appl
12	98.4	40.0	4689	9	US-09-988-384B-34	Sequence 34, Appl
13	98.4	40.0	4689	9	US-09-836-821-34	Sequence 34, Appl
14	98.4	40.0	4689	10	US-09-861-289-34	Sequence 34, Appl
15	96.2	39.1	15872	9	US-09-860-846-1	Sequence 1, Appli
16	96.2	39.1	15872	9	US-09-988-384B-1	Sequence 1, Appli
17	96.2	39.1	15872	9	US-09-836-821-1	Sequence 1, Appli
18	96.2	39.1	15872	10	US-09-861-289-1	Sequence 1, Appli
19	94.6	38.5	13842	9	US-09-860-846-30	Sequence 30, Appl
20	94.6	38.5	13842	9	US-09-988-384B-30	Sequence 30, Appl
21	94.6	38.5	13842	9	US-09-836-821-30	Sequence 30, Appl
22	94.6	38.5	13842	10	US-09-861-289-30	Sequence 30, Appl
23	86	35.0	4209	9	US-09-712-363-20	Sequence 20, Appl
24	80.8	32.8	4041	9	US-09-860-846-36	Sequence 36, Appl
25	80.8	32.8	4041	9	US-09-988-384B-36	Sequence 36, Appl
26	80.8	32.8	4041	9	US-09-836-821-36	Sequence 36, Appl
27	80.8	32.8	4041	10	US-09-861-289-36	Sequence 36, Appl
28	72.2	29.3	4851	9	US-09-712-363-116	Sequence 116, App
29	59	24.0	68750	9	US-10-014-717-1	Sequence 1, Appli
30	44.4	18.0	88421	9	US-09-976-059-1	Sequence 1, Appli
31	33.6	13.7	2305	10	US-09-073-009-2	Sequence 2, Appli
32	33.6	13.7	2305	10	US-09-023-588-2	Sequence 2, Appli
33	33.6	13.7	2305	10	US-09-793-306-2	Sequence 2, Appli
34	33.4	13.6	1134	9	US-09-738-626-2238	Sequence 2238, Ap
35	33.4	13.6	2168	9	US-10-198-846-12967	Sequence 12967, A
36	33.4	13.6	3309400	9	US-09-738-626-1	Sequence 1, Appli
c 37	32.6	13.3	1471	9	US-09-992-598-496	Sequence 496, App
c 38	32.6	13.3	1471	9	US-09-989-293A-496	Sequence 496, App
c 39	32.6	13.3	1471	9	US-09-989-735-496	Sequence 496, App
c 40	32.6	13.3	1471	9	US-09-990-444-496	Sequence 496, App
c 41	32.6	13.3	1471	9	US-09-989-730-496	Sequence 496, App
c 42	32.6	13.3	1471	9	US-09-990-436-496	Sequence 496, App
c 43	32.6	13.3	1471	9	US-09-991-181-496	Sequence 496, App
c 44	32.6	13.3	1471	9	US-09-993-687-496	Sequence 496, App
c 45	32.6	13.3	1471	9	US-09-989-734-496	Sequence 496, App

ALIGNMENTS

RESULT 1
US-09-808-880-1
; Sequence 1, Application US/09808880
; Publication No. US20030027287A1
; GENERAL INFORMATION:
; APPLICANT: Betlach, Mary C.
; APPLICANT: Shah, Sanjay Krishnakant
; APPLICANT: McDaniel, Robert
; APPLICANT: Tang, Li
; TITLE OF INVENTION: RECOMBINANT OLEANDOLIDE POLYKETIDE SYNTHASE
; FILE REFERENCE: 30062-20029.00
; CURRENT APPLICATION NUMBER: US/09/808,880
; CURRENT FILING DATE: 2001-03-14
; PRIOR APPLICATION NUMBER: US/09/428,517
; PRIOR FILING DATE: 1999-10-28
; PRIOR APPLICATION NUMBER: 60/120,254
; PRIOR FILING DATE: 1999-02-16
; PRIOR APPLICATION NUMBER: 60/106,100
; PRIOR FILING DATE: 1998-10-29
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 50937
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artifical Sequence: Recombinant DNA
US-09-808-880-1

Query Match 44.8%; Score 110.2; DB 9; Length 50937;
Best Local Similarity 66.8%; Pred. No. 1.5e-25;
Matches 157; Conservative 0; Mismatches 78; Indels 0; Gaps 0;

```
Qy          12 GACGTCGGCCGACCAACGGGCCGCATTGCTCGAGCTCGTACGAGACCATGTGGCGGCAGT 71
               |||| |  | |||| |||     ||  ||| ||| ||||| || | ||| || ||    ||
Db       29150 GGCGGCACCGGACCCACGTCGTGCCCTGCGCGACCTCGTCCGCGGCCACGTCGCCCTGGT 29209

Qy          72 GCTCCGGCACGCGGACCCGAAAGCCATCGCGCCCGACCAGTCGTTCCGTGCACTCGGCTT 131
               |||| | |||   | |||  | |||||||   |||| ||| | ||||| |   |||| ||
Db       29210 GCTCGGACACAGTGGCCCCGAGGCCATCGACGCCGAACAGGCCTTCCGGGACATCGGTTT 29269

Qy         132 CGATTCACTCACGGCCGTCGAGTTCCGAAACCTGCTGATCAAGGCAACAGGACTCCGCCT 191
```

```
                ||| || || || || |||||  || ||||||  |||||  |    |  || || ||||||  |
Db      29270 CGACTCCCTGACCGCAGTCGAACTCAGAAACCGGCTGAACGCCGAGACCGGCCTCCGCTT 29329

Qy        192 TCCTGTCTCGCTGGTCTTCGACCACCCGACCCCTGCCAAACTCGCCGTACACCTG 246
              || | | |||| || ||||||  |||| | |||    |  |||||||  ||||||
Db      29330 GCCCGGCACGCTCGTGTTCGACTACCCCAACCCGAGCGCGCTCGCCGATCACCTG 29384


RESULT 2
US-09-860-846-32
; Sequence 32, Application US/09860846
; Patent No. US20020164742A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/860,846
;  CURRENT FILING DATE:  2001-05-18
;  PRIOR APPLICATION NUMBER: 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 32
;   LENGTH: 11220
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-860-846-32

  Query Match             42.1%;  Score 103.6;  DB 9;  Length 11220;
  Best Local Similarity   63.8%;  Pred. No. 1.7e-23;
  Matches  157;  Conservative    0;  Mismatches   89;  Indels    0;  Gaps    0;

Qy          1 CTGGCCCATCAGACGTCGGCCGACCAACGGGCCGCATTGCTCGAGCTCGTACGAGACCAT 60
              || |||      |||  ||| |||||      ||      |||  || ||||| ||    ||
Db      10678 CTCGCCGCGATGACACCGGACGACCGGGTCGCGCACCTGCGGGACCTCGTCCGTACGCAC 10737

Qy         61 GTGGCGGCAGTGCTCCGGCACGCGGACCCGAAAGCCATCGCGCCCGACCAGTCGTTCCGT 120
              |||||| | || ||  | ||||     |||||        | |  |  || | | | |||||
Db      10738 GTGGCGACCGTCCTGGGACACGGCACCCCGAGCCGGGTGGACCTGGAGCGGGCCTTCCGC 10797

Qy        121 GCACTCGGCTTCGATTCACTCACGGCCGTCGAGTTCCGAAACCTGCTGATCAAGGCAACA 180
              |   ||| ||||| || ||||| ||||||||  |||| ||||  || | |   || ||
Db      10798 GACACCGGTTTCGACTCGCTCACCGCCGTCGAACTCCGCAACCGTCTCAACGCCGCGACC 10857

Qy        181 GGACTCCGCCTTCCTGTCTCGCTGGTCTTCGACCACCCGACCCCTGCCAAACTCGCCGTA 240
              || || || || || | | ||||||||||||||||||| ||||| |   | |||||||
Db      10858 GGGCTGCGGCTGCCGGCCACGCTGGTCTTCGACCACCCCACCCCGGGGGAGCTCGCCGGG 10917

Qy        241 CACCTG 246
              ||||||
Db      10918 CACCTG 10923


RESULT 3
US-09-988-384B-32
; Sequence 32, Application US/09988384B
; Publication No. US20030073824A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.536US1
;  CURRENT APPLICATION NUMBER: US/09/988,384B
;  CURRENT FILING DATE:  2001-11-19
;  PRIOR APPLICATION NUMBER: PCT/US99/14398
;  PRIOR FILING DATE: 1999-06-25
;  PRIOR APPLICATION NUMBER: US 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 53
; SEQ ID NO 32
;   LENGTH: 11220
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-988-384B-32

  Query Match             42.1%;  Score 103.6;  DB 9;  Length 11220;
  Best Local Similarity   63.8%;  Pred. No. 1.7e-23;
  Matches  157;  Conservative    0;  Mismatches   89;  Indels    0;  Gaps    0;

Qy          1 CTGGCCCATCAGACGTCGGCCGACCAACGGGCCGCATTGCTCGAGCTCGTACGAGACCAT 60
              || |||      |||  ||| |||||      ||      |||  || ||||| ||    ||
Db      10678 CTCGCCGCGATGACACCGGACGACCGGGTCGCGCACCTGCGGGACCTCGTCCGTACGCAC 10737

Qy         61 GTGGCGGCAGTGCTCCGGCACGCGGACCCGAAAGCCATCGCGCCCGACCAGTCGTTCCGT 120
              |||||| | || ||  | ||||     |||||        | |  |  || | | | |||||
Db      10738 GTGGCGACCGTCCTGGGACACGGCACCCCGAGCCGGGTGGACCTGGAGCGGGCCTTCCGC 10797

Qy        121 GCACTCGGCTTCGATTCACTCACGGCCGTCGAGTTCCGAAACCTGCTGATCAAGGCAACA 180
              |   ||| ||||| || ||||| ||||||||  |||| ||||  || | |   || ||
Db      10798 GACACCGGTTTCGACTCGCTCACCGCCGTCGAACTCCGCAACCGTCTCAACGCCGCGACC 10857

Qy        181 GGACTCCGCCTTCCTGTCTCGCTGGTCTTCGACCACCCGACCCCTGCCAAACTCGCCGTA 240
              || || || || || | | ||||||||||||||||||| ||||| |   | |||||||
Db      10858 GGGCTGCGGCTGCCGGCCACGCTGGTCTTCGACCACCCCACCCCGGGGGAGCTCGCCGGG 10917

Qy        241 CACCTG 246
              ||||||
Db      10918 CACCTG 10923


RESULT 4
US-09-836-821-32
; Sequence 32, Application US/09836821
; Publication No. US20030087405A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/836,821
;  CURRENT FILING DATE:  2001-04-17
;  PRIOR APPLICATION NUMBER: 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 32
;   LENGTH: 11220
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-836-821-32

  Query Match             42.1%;  Score 103.6;  DB 9;  Length 11220;
  Best Local Similarity   63.8%;  Pred. No. 1.7e-23;
  Matches  157;  Conservative    0;  Mismatches   89;  Indels    0;  Gaps    0;

Qy          1 CTGGCCCATCAGACGTCGGCCGACCAACGGGCCGCATTGCTCGAGCTCGTACGAGACCAT 60
              || |||      |||  ||| |||||      ||      |||  || ||||| ||    ||
Db      10678 CTCGCCGCGATGACACCGGACGACCGGGTCGCGCACCTGCGGGACCTCGTCCGTACGCAC 10737

Qy         61 GTGGCGGCAGTGCTCCGGCACGCGGACCCGAAAGCCATCGCGCCCGACCAGTCGTTCCGT 120
              |||||| | || ||  | ||||     |||||        | |  |  || | | | |||||
Db      10738 GTGGCGACCGTCCTGGGACACGGCACCCCGAGCCGGGTGGACCTGGAGCGGGCCTTCCGC 10797

Qy        121 GCACTCGGCTTCGATTCACTCACGGCCGTCGAGTTCCGAAACCTGCTGATCAAGGCAACA 180
              |   ||| ||||| || ||||| ||||||||  |||| ||||  || | |   || ||
Db      10798 GACACCGGTTTCGACTCGCTCACCGCCGTCGAACTCCGCAACCGTCTCAACGCCGCGACC 10857

Qy        181 GGACTCCGCCTTCCTGTCTCGCTGGTCTTCGACCACCCGACCCCTGCCAAACTCGCCGTA 240
              || || || || || | | ||||||||||||||||||| ||||| |   | |||||||
```

```
Db      10858 GGGCTGCGGCTGCCGGCCACGCTGGTCTTCGACCACCCCACCCCGGGGGAGCTCGCCGGG 10917

Qy        241 CACCTG 246
              ||||||
Db      10918 CACCTG 10923


RESULT 5
US-09-861-289-32
; Sequence 32, Application US/09861289
; Patent No. US20020110897A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/861,289
;  CURRENT FILING DATE:  2001-05-18
;  PRIOR APPLICATION NUMBER: 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 32
;   LENGTH: 11220
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-861-289-32

  Query Match             42.1%;  Score 103.6;  DB 10;  Length 11220;
  Best Local Similarity   63.8%;  Pred. No. 1.7e-23;
  Matches  157;  Conservative    0;  Mismatches   89;  Indels    0;  Gaps    0;

Qy          1 CTGGCCCATCAGACGTCGGCCGACCAACGGGCCGCATTGCTCGAGCTCGTACGAGACCAT 60
              || |||      |||  ||| |||||      ||      |||  || ||||| ||    ||
Db      10678 CTCGCCGCGATGACACCGGACGACCGGGTCGCGCACCTGCGGGACCTCGTCCGTACGCAC 10737

Qy         61 GTGGCGGCAGTGCTCCGGCACGCGGACCCGAAAGCCATCGCGCCCGACCAGTCGTTCCGT 120
              |||||| | || ||  | ||||     |||||       | |  |  || | | | |||||
Db      10738 GTGGCGACCGTCCTGGGACACGGCACCCCGAGCCGGGTGGACCTGGAGCGGGCCTTCCGC 10797

Qy        121 GCACTCGGCTTCGATTCACTCACGGCCGTCGAGTTCCGAAACCTGCTGATCAAGGCAACA 180
              |    ||| ||||| || ||||| ||||||||  |||| ||||  || | |   || ||
Db      10798 GACACCGGTTTCGACTCGCTCACCGCCGTCGAACTCCGCAACCGTCTCAACGCCGCGACC 10857

Qy        181 GGACTCCGCCTTCCTGTCTCGCTGGTCTTCGACCACCCGACCCCTGCCAAACTCGCCGTA 240
              || || || || || | | ||||||||||||||||||| ||||| |   | |||||||
Db      10858 GGGCTGCGGCTGCCGGCCACGCTGGTCTTCGACCACCCCACCCCGGGGGAGCTCGCCGGG 10917

Qy        241 CACCTG 246
              ||||||
Db      10918 CACCTG 10923


RESULT 6
US-09-860-846-5
; Sequence 5, Application US/09860846
; Patent No. US20020164742A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/860,846
;  CURRENT FILING DATE:  2001-05-18
;  PRIOR APPLICATION NUMBER: 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 5
;   LENGTH: 36778
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-860-846-5

  Query Match             42.1%;  Score 103.6;  DB 9;  Length 36778;
  Best Local Similarity   63.8%;  Pred. No. 1.9e-23;
  Matches  157;  Conservative    0;  Mismatches   89;  Indels    0;  Gaps    0;

Qy          1 CTGGCCCATCAGACGTCGGCCGACCAACGGGCCGCATTGCTCGAGCTCGTACGAGACCAT 60
              || |||      |||  ||| |||||      ||      |||  || ||||| ||    ||
Db      26365 CTCGCCGCGATGACACCGGACGACCGGGTCGCGCACCTGCGGGACCTCGTCCGTACGCAC 26424

Qy         61 GTGGCGGCAGTGCTCCGGCACGCGGACCCGAAAGCCATCGCGCCCGACCAGTCGTTCCGT 120
              |||||| | || ||  | ||||     |||||       | |  |  || | | | |||||
Db      26425 GTGGCGACCGTCCTGGGACACGGCACCCCGAGCCGGGTGGACCTGGAGCGGGCCTTCCGC 26484

Qy        121 GCACTCGGCTTCGATTCACTCACGGCCGTCGAGTTCCGAAACCTGCTGATCAAGGCAACA 180
              |    ||| ||||| || ||||| ||||||||  |||| ||||  || | |   || ||
Db      26485 GACACCGGTTTCGACTCGCTCACCGCCGTCGAACTCCGCAACCGTCTCAACGCCGCGACC 26544

Qy        181 GGACTCCGCCTTCCTGTCTCGCTGGTCTTCGACCACCCGACCCCTGCCAAACTCGCCGTA 240
              || || || || || | | ||||||||||||||||||| ||||| |   | |||||||
Db      26545 GGGCTGCGGCTGCCGGCCACGCTGGTCTTCGACCACCCCACCCCGGGGGAGCTCGCCGGG 26604

Qy        241 CACCTG 246
              ||||||
Db      26605 CACCTG 26610


RESULT 7
US-09-836-821-5
; Sequence 5, Application US/09836821
; Publication No. US20030087405A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/836,821
;  CURRENT FILING DATE:  2001-04-17
;  PRIOR APPLICATION NUMBER: 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 5
;   LENGTH: 36778
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-836-821-5

  Query Match             42.1%;  Score 103.6;  DB 9;  Length 36778;
  Best Local Similarity   63.8%;  Pred. No. 1.9e-23;
  Matches  157;  Conservative    0;  Mismatches   89;  Indels    0;  Gaps    0;

Qy          1 CTGGCCCATCAGACGTCGGCCGACCAACGGGCCGCATTGCTCGAGCTCGTACGAGACCAT 60
              || |||      |||  ||| |||||      ||      |||  || ||||| ||    ||
Db      26365 CTCGCCGCGATGACACCGGACGACCGGGTCGCGCACCTGCGGGACCTCGTCCGTACGCAC 26424

Qy         61 GTGGCGGCAGTGCTCCGGCACGCGGACCCGAAAGCCATCGCGCCCGACCAGTCGTTCCGT 120
              |||||| | || ||  | ||||     |||||       | |  |  || | | | |||||
Db      26425 GTGGCGACCGTCCTGGGACACGGCACCCCGAGCCGGGTGGACCTGGAGCGGGCCTTCCGC 26484

Qy        121 GCACTCGGCTTCGATTCACTCACGGCCGTCGAGTTCCGAAACCTGCTGATCAAGGCAACA 180
              |    ||| ||||| || ||||| ||||||||  |||| ||||  || | |   || ||
Db      26485 GACACCGGTTTCGACTCGCTCACCGCCGTCGAACTCCGCAACCGTCTCAACGCCGCGACC 26544

Qy        181 GGACTCCGCCTTCCTGTCTCGCTGGTCTTCGACCACCCGACCCCTGCCAAACTCGCCGTA 240
              || || || || || | | ||||||||||||||||||| ||||| |   | |||||||
Db      26545 GGGCTGCGGCTGCCGGCCACGCTGGTCTTCGACCACCCCACCCCGGGGGAGCTCGCCGGG 26604
```

```
Qy        241 CACCTG 246
              ||||||
Db      26605 CACCTG 26610


RESULT 8
US-09-861-289-5
; Sequence 5, Application US/09861289
; Patent No. US20020110897A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/861,289
;  CURRENT FILING DATE:  2001-05-18
;  PRIOR APPLICATION NUMBER: 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 5
;   LENGTH: 36778
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-861-289-5

  Query Match             42.1%;  Score 103.6;  DB 10;  Length 36778;
  Best Local Similarity   63.8%;  Pred. No. 1.9e-23;
  Matches  157;  Conservative    0;  Mismatches   89;  Indels    0;  Gaps    0;

Qy          1 CTGGCCCATCAGACGTCGGCCGACCAACGGGCCGCATTGCTCGAGCTCGTACGAGACCAT 60
              || |||     |||  ||| |||||     ||      |||  || ||||| ||    ||
Db      26365 CTCGCCGCGATGACACCGGACGACCGGGTCGCGCACCTGCGGGACCTCGTCCGTACGCAC 26424

Qy         61 GTGGCGGCAGTGCTCCGGCACGCGGACCCGAAAGCCATCGCGCCCGACCAGTCGTTCCGT 120
              ||||||  | ||  ||  | ||||    |||||       | |  |  || | | | |||||
Db      26425 GTGGCGACCGTCCTGGGACACGGCACCCCGAGCCGGGTGGACCTGGAGCGGGCCTTCCGC 26484

Qy        121 GCACTCGGCTTCGATTCACTCACGGCCGTCGAGTTCCGAAACCTGCTGATCAAGGCAACA 180
              |    ||| ||||| || ||||| ||||||||  |||| ||||  || | |    || ||
Db      26485 GACACCGGTTTCGACTCGCTCACCGCCGTCGAACTCCGCAACCGTCTCAACGCCGCGACC 26544

Qy        181 GGACTCCGCCTTCCTGTCTCGCTGGTCTTCGACCACCCGACCCCTGCCAAACTCGCCGTA 240
              || || || ||  || | | ||||||||||||||||||| |||||  |   | |||||||
Db      26545 GGGCTGCGGCTGCCGGCCACGCTGGTCTTCGACCACCCCACCCCGGGGGAGCTCGCCGGG 26604

Qy        241 CACCTG 246
              ||||||
Db      26605 CACCTG 26610


RESULT 9
US-09-988-384B-5
; Sequence 5, Application US/09988384B
; Publication No. US20030073824A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.536US1
;  CURRENT APPLICATION NUMBER: US/09/988,384B
;  CURRENT FILING DATE:  2001-11-19
;  PRIOR APPLICATION NUMBER: PCT/US99/14398
;  PRIOR FILING DATE: 1999-06-25
;  PRIOR APPLICATION NUMBER: US 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 53
; SEQ ID NO 5
;   LENGTH: 37948
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-988-384B-5

  Query Match             42.1%;  Score 103.6;  DB 9;  Length 37948;
  Best Local Similarity   63.8%;  Pred. No. 1.9e-23;
  Matches  157;  Conservative    0;  Mismatches   89;  Indels    0;  Gaps    0;

Qy          1 CTGGCCCATCAGACGTCGGCCGACCAACGGGCCGCATTGCTCGAGCTCGTACGAGACCAT 60
              || |||     |||  ||| |||||     ||      |||  || ||||| ||    ||
Db      27535 CTCGCCGCGATGACACCGGACGACCGGGTCGCGCACCTGCGGGACCTCGTCCGTACGCAC 27594

Qy         61 GTGGCGGCAGTGCTCCGGCACGCGGACCCGAAAGCCATCGCGCCCGACCAGTCGTTCCGT 120
              ||||||  | ||  ||  | ||||    |||||       | |  |  || | | | |||||
Db      27595 GTGGCGACCGTCCTGGGACACGGCACCCCGAGCCGGGTGGACCTGGAGCGGGCCTTCCGC 27654

Qy        121 GCACTCGGCTTCGATTCACTCACGGCCGTCGAGTTCCGAAACCTGCTGATCAAGGCAACA 180
              |    ||| ||||| || ||||| ||||||||  |||| ||||  || | |    || ||
Db      27655 GACACCGGTTTCGACTCGCTCACCGCCGTCGAACTCCGCAACCGTCTCAACGCCGCGACC 27714

Qy        181 GGACTCCGCCTTCCTGTCTCGCTGGTCTTCGACCACCCGACCCCTGCCAAACTCGCCGTA 240
              || || || ||  || | | ||||||||||||||||||| |||||  |   | |||||||
Db      27715 GGGCTGCGGCTGCCGGCCACGCTGGTCTTCGACCACCCCACCCCGGGGGAGCTCGCCGGG 27774

Qy        241 CACCTG 246
              ||||||
Db      27775 CACCTG 27780


RESULT 10
US-09-793-708-19
; Sequence 19, Application US/09793708
; Publication No. US20030104597A1
; GENERAL INFORMATION:
;  APPLICANT: ASHLEY, Gary
;  APPLICANT:  BETLACH, Melanie C.
;  APPLICANT:  BETLACH, Mary C.
;  APPLICANT:  McDANIEL, Robert
;  APPLICANT:  TANG, Li
;  TITLE OF INVENTION: RECOMBINANT NARBONOLIDE POLYKETIDE SYNTHASE
;  FILE REFERENCE: 300622002121
;  CURRENT APPLICATION NUMBER: US/09/793,708
;  CURRENT FILING DATE:  2001-02-22
;  PRIOR APPLICATION NUMBER: US 09/657,440
;  PRIOR FILING DATE: 2000-09-07
;  PRIOR APPLICATION NUMBER: US 09/320,878
;  PRIOR FILING DATE: 1999-05-27
;  PRIOR APPLICATION NUMBER: US 09/141,908
;  PRIOR FILING DATE: 1998-08-28
;  PRIOR APPLICATION NUMBER: US 09/073,538
;  PRIOR FILING DATE: 1998-05-06
;  PRIOR APPLICATION NUMBER: US 08/846,247
;  PRIOR FILING DATE: 1997-04-30
;  PRIOR APPLICATION NUMBER: US 60/134,990
;  PRIOR FILING DATE: 1999-05-20
;  NUMBER OF SEQ ID NOS: 38
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 19
;   LENGTH: 38506
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-793-708-19

  Query Match             42.1%;  Score 103.6;  DB 9;  Length 38506;
  Best Local Similarity   63.8%;  Pred. No. 1.9e-23;
  Matches  157;  Conservative    0;  Mismatches   89;  Indels    0;  Gaps    0;

Qy          1 CTGGCCCATCAGACGTCGGCCGACCAACGGGCCGCATTGCTCGAGCTCGTACGAGACCAT 60
              || |||     |||  ||| |||||     ||      |||  || ||||| ||    ||
Db      24507 CTCGCCGCGATGACACCGGACGACCGGGTCGCGCACCTGCGGGACCTCGTCCGTACGCAC 24566

Qy         61 GTGGCGGCAGTGCTCCGGCACGCGGACCCGAAAGCCATCGCGCCCGACCAGTCGTTCCGT 120
              ||||||  | ||  ||  | ||||    |||||       | |  |  || | | | |||||
```

```
Db      24567 GTGGCGACCGTCCTGGGACACGGCACCCCGAGCCGGGTGGACCTGGAGCGGGCCTTCCGC 24626

Qy        121 GCACTCGGCTTCGATTCACTCACGGCCGTCGAGTTCCGAAACCTGCTGATCAAGGCAACA 180
              |    ||| ||||| || ||||| |||||||| |||| ||||  || | |    || ||
Db      24627 GACACCGGTTTCGACTCGCTCACCGCCGTCGAACTCCGCAACCGTCTCAACGCCGCGACC 24686

Qy        181 GGACTCCGCCTTCCTGTCTCGCTGGTCTTCGACCACCCGACCCCTGCCAAACTCGCCGTA 240
              || || || || || | | ||||||||||||||||||| ||||| |   | |||||||
Db      24687 GGGCTGCGGCTGCCGGCCACGCTGGTCTTCGACCACCCCACCCCGGGGGAGCTCGCCGGG 24746

Qy        241 CACCTG 246
              ||||||
Db      24747 CACCTG 24752


RESULT 11
US-09-860-846-34
; Sequence 34, Application US/09860846
; Patent No. US20020164742A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/860,846
;  CURRENT FILING DATE:  2001-05-18
;  PRIOR APPLICATION NUMBER: 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 34
;   LENGTH: 4689
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-860-846-34

  Query Match             40.0%;  Score 98.4;  DB 9;  Length 4689;
  Best Local Similarity   64.5%;  Pred. No. 7.3e-22;
  Matches  147;  Conservative    0;  Mismatches   81;  Indels    0;  Gaps    0;

Qy         18 GGCCGACCAACGGGCCGCATTGCTCGAGCTCGTACGAGACCATGTGGCGGCAGTGCTCCG 77
              | |||| |  ||| | ||  | |||   ||||| ||   ||| | |||||| || ||| |
Db       4194 GCCCGAGCGCCGGCCGGCGCTCCTCACCCTCGTCCGTACCCACGCGGCGGCCGTACTCGG 4253

Qy         78 GCACGCGGACCCGAAAGCCATCGCGCCCGACCAGTCGTTCCGTGCACTCGGCTTCGATTC 137
              ||  |   |||  |     | || |||| ||   | |||   |  ||||||||||| ||
Db       4254 CCATTCCTCCCCCGACCGGGTGGCCCCCGGCCGTGCCTTCACCGAGCTCGGCTTCGACTC 4313

Qy        138 ACTCACGGCCGTCGAGTTCCGAAACCTGCTGATCAAGGCAACAGGACTCCGCCTTCCTGT 197
              || ||||||||  || |||| |||| |||   || ||      ||   | | || || |
Db       4314 GCTGACGGCCGTCGCAGCTCCGCAACCAGCTCTCCACGGTGGTCGGCAACAGGCTCCCCGC 4373

Qy        198 CTCGCTGGTCTTCGACCACCCGACCCCTGCCAAACTCGCCGTACACCT 245
              | |   |||||||||||||||||| || |||  ||||||||  ||||
Db       4374 CACCACGGTCTTCGACCACCCGACGCCCGCCGCACTCGCCGCGCACCT 4421


RESULT 12
US-09-988-384B-34
; Sequence 34, Application US/09988384B
; Publication No. US20030073824A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.536US1
;  CURRENT APPLICATION NUMBER: US/09/988,384B
;  CURRENT FILING DATE:  2001-11-19
;  PRIOR APPLICATION NUMBER: PCT/US99/14398
;  PRIOR FILING DATE: 1999-06-25
;  PRIOR APPLICATION NUMBER: US 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 53
; SEQ ID NO 34
;   LENGTH: 4689
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-988-384B-34

  Query Match             40.0%;  Score 98.4;  DB 9;  Length 4689;
  Best Local Similarity   64.5%;  Pred. No. 7.3e-22;
  Matches  147;  Conservative    0;  Mismatches   81;  Indels    0;  Gaps    0;

Qy         18 GGCCGACCAACGGGCCGCATTGCTCGAGCTCGTACGAGACCATGTGGCGGCAGTGCTCCG 77
              | |||| |  ||| | ||  | |||   ||||| ||   ||| | |||||| || ||| |
Db       4194 GCCCGAGCGCCGGCCGGCGCTCCTCACCCTCGTCCGTACCCACGCGGCGGCCGTACTCGG 4253

Qy         78 GCACGCGGACCCGAAAGCCATCGCGCCCGACCAGTCGTTCCGTGCACTCGGCTTCGATTC 137
              ||  |   |||  |     | || |||| ||   | |||   |  ||||||||||| ||
Db       4254 CCATTCCTCCCCCGACCGGGTGGCCCCCGGCCGTGCCTTCACCGAGCTCGGCTTCGACTC 4313

Qy        138 ACTCACGGCCGTCGAGTTCCGAAACCTGCTGATCAAGGCAACAGGACTCCGCCTTCCTGT 197
              || ||||||||  || |||| |||| |||   || ||      ||   | | || || |
Db       4314 GCTGACGGCCGTCGCAGCTCCGCAACCAGCTCTCCACGGTGGTCGGCAACAGGCTCCCCGC 4373

Qy        198 CTCGCTGGTCTTCGACCACCCGACCCCTGCCAAACTCGCCGTACACCT 245
              | |   |||||||||||||||||| || |||  ||||||||  ||||
Db       4374 CACCACGGTCTTCGACCACCCGACGCCCGCCGCACTCGCCGCGCACCT 4421


RESULT 13
US-09-836-821-34
; Sequence 34, Application US/09836821
; Publication No. US20030087405A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/836,821
;  CURRENT FILING DATE:  2001-04-17
;  PRIOR APPLICATION NUMBER: 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 34
;   LENGTH: 4689
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-836-821-34

  Query Match             40.0%;  Score 98.4;  DB 9;  Length 4689;
  Best Local Similarity   64.5%;  Pred. No. 7.3e-22;
  Matches  147;  Conservative    0;  Mismatches   81;  Indels    0;  Gaps    0;

Qy         18 GGCCGACCAACGGGCCGCATTGCTCGAGCTCGTACGAGACCATGTGGCGGCAGTGCTCCG 77
              | |||| |  ||| | ||  | |||   ||||| ||   ||| | |||||| || ||| |
Db       4194 GCCCGAGCGCCGGCCGGCGCTCCTCACCCTCGTCCGTACCCACGCGGCGGCCGTACTCGG 4253

Qy         78 GCACGCGGACCCGAAAGCCATCGCGCCCGACCAGTCGTTCCGTGCACTCGGCTTCGATTC 137
              ||  |   |||  |     | || |||| ||   | |||   |  ||||||||||| ||
Db       4254 CCATTCCTCCCCCGACCGGGTGGCCCCCGGCCGTGCCTTCACCGAGCTCGGCTTCGACTC 4313

Qy        138 ACTCACGGCCGTCGAGTTCCGAAACCTGCTGATCAAGGCAACAGGACTCCGCCTTCCTGT 197
              || ||||||||  || |||| |||| |||   || ||      ||   | | || || |
Db       4314 GCTGACGGCCGTCGCAGCTCCGCAACCAGCTCTCCACGGTGGTCGGCAACAGGCTCCCCGC 4373

Qy        198 CTCGCTGGTCTTCGACCACCCGACCCCTGCCAAACTCGCCGTACACCT 245
              | |   |||||||||||||||||| || |||  ||||||||  ||||
Db       4374 CACCACGGTCTTCGACCACCCGACGCCCGCCGCACTCGCCGCGCACCT 4421
```

```
RESULT 14
US-09-861-289-34
; Sequence 34, Application US/09861289
; Patent No. US20020110897A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/861,289
;  CURRENT FILING DATE:  2001-05-18
;  PRIOR APPLICATION NUMBER: 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 34
;   LENGTH: 4689
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-861-289-34

  Query Match             40.0%;  Score 98.4;  DB 10;  Length 4689;
  Best Local Similarity   64.5%;  Pred. No. 7.3e-22;
  Matches  147;  Conservative    0;  Mismatches   81;  Indels    0;  Gaps    0;

Qy          18 GGCCGACCAACGGGCCGCATTGCTCGAGCTCGTACGAGACCATGTGGCGGCAGTGCTCCG 77
               | |||| |  ||| | ||  | |||   ||||| ||   ||| | |||||| || ||| |
Db        4194 GCCCGAGCGCCGGCCGGCGCTCCTCACCCTCGTCCGTACCCACGCGGCGGCCGTACTCGG 4253

Qy          78 GCACGCGGACCCGAAAGCCATCGCGCCCGACCAGTCGTTCCGTGCACTCGGCTTCGATTC 137
               ||  |   |||  |      | || |||| ||   | |||   |  |||||||||| ||
Db        4254 CCATTCCTCCCCCGACCGGGTGGCCCCCGGCCGTGCCTTCACCGAGCTCGGCTTCGACTC 4313

Qy         138 ACTCACGGCCGTCGAGTTCCGAAACCTGCTGATCAAGGCAACAGGACTCCGCCTTCCTGT 197
               || ||||||||  || |||| |||| |||   || ||      ||   | | || || |
Db        4314 GCTGACGGCCGTGCAGCTCCGCAACCAGCTCTCCACGGTGGTCGGCAACAGGCTCCCCGC 4373

Qy         198 CTCGCTGGTCTTCGACCACCCGACCCCTGCCAAACTCGCCGTACACCT 245
               | |   |||||||||||||||||| || |||  ||||||||  |||||
Db        4374 CACCACGGTCTTCGACCACCCGACGCCCGCCGCACTCGCCGCGCACCT 4421


RESULT 15
US-09-860-846-1
; Sequence 1, Application US/09860846
; Patent No. US20020164742A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/860,846
;  CURRENT FILING DATE:  2001-05-18
;  PRIOR APPLICATION NUMBER: 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 1
;   LENGTH: 15872
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-860-846-1

  Query Match             39.1%;  Score 96.2;  DB 9;  Length 15872;
  Best Local Similarity   66.8%;  Pred. No. 3.9e-21;
  Matches  137;  Conservative    0;  Mismatches   68;  Indels    0;  Gaps    0;

Qy          42 CGAGCTCGTACGAGACCATGTGGCGGCAGTGCTCCGGCACGCGGACCCGAAAGCCATCGC 101
               ||||| ||| || |  || || || || || ||| | |||  |||||||     | |||
Db        2815 CGAGCGCGTGCGCGCACACGTCGCCGCCGTCCTCGAGTACGACGACCCGACCCGCGTCGA 2874

Qy         102 GCCCGACCAGTCGTTCCGTGCACTCGGCTTCGATTCACTCACGGCCGTCGAGTTCCGAAA 161
               | || ||   | |||   |  || |||||||| || |||| | |||||||| | || ||
Db        2875 ACTCGGCCTCACCTTCAAGGAGCTGGGCTTCGACTCCCTCATGTCCGTCGAGCTGCGGAA 2934

Qy         162 CCTGCTGATCAAGGCAACAGGACTCCGCCTTCCTGTCTCGCTGGTCTTCGACCACCCGAC 221
               |  |||  || | |  || ||||| ||||| ||   |   ||| ||||||||||||||||
Db        2935 CGCGCTCGTCGACGACACGGGACTGCGCCTGCCCAGCGGACTGCTCTTCGACCACCCGAC 2994

Qy         222 CCCTGCCAAACTCGCCGTACACCTG 246
               ||   |   |||||||  ||||||
Db        2995 GCCGCGCGCCCTCGCCGCCCACCTG 3019


Search completed: June 18, 2003, 05:08:27
Job time : 32.1628 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 17, 2003, 11:26:45 ; Search time 315.347 Seconds
(without alignments)
12633.968 Million cell updates/sec

Title: US-09-914-286-1_COPY_5935_6180
Perfect score: 246
Sequence: 1 ctggcccatcagacgtcggc.....ccaaactcgccgtacacctg 246

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	45	18.3	668	13	BM629517	BM629517 170006686
2	45	18.3	672	13	BM620935	BM620935 170006874
3	45	18.3	687	13	BM633427	BM633427 170006875
4	45	18.3	698	13	BM650789	BM650789 170006873
5	45	18.3	715	13	BM653397	BM653397 170006873
6	43.4	17.6	509	13	BM638546	BM638546 170006875
7	43.4	17.6	549	13	BM609994	BM609994 170006871
8	43.4	17.6	567	13	BM649092	BM649092 170006873
9	43.4	17.6	602	13	BM632295	BM632295 170006875
10	43.4	17.6	639	13	BM632995	BM632995 170006875
11	43.4	17.6	663	13	BM645157	BM645157 170006873
12	43.4	17.6	678	13	BM627481	BM627481 170006874
13	43.4	17.6	678	13	BM656823	BM656823 170006873
14	43.4	17.6	684	13	BM634614	BM634614 170006875
15	43.4	17.6	684	13	BM648344	BM648344 170006873
16	43.4	17.6	693	13	BM647510	BM647510 170006873
17	43.4	17.6	699	13	BM635786	BM635786 170006875
18	43.4	17.6	703	13	BM628514	BM628514 170006874
19	43.4	17.6	721	13	BM653415	BM653415 170006873
20	40.2	16.3	637	13	BM623435	BM623435 170006874
21	40.2	16.3	637	13	BM639126	BM639126 170006875
c 22	37	15.0	525	13	BI784654	BI784654 saf57f06.
23	37	15.0	607	14	BQ811315	BQ811315 1030022A0
24	37	15.0	925	17	CNS0091P	AL053013 Drosophil
25	36.6	14.9	435	9	AJ455524	AJ455524 AJ455524
c 26	36.4	14.8	459	9	AL818182	AL818182 AL818182
c 27	36.4	14.8	477	13	BJ234991	BJ234991 BJ234991
c 28	36.4	14.8	487	14	BQ244817	BQ244817 TaE15035B
c 29	36.4	14.8	509	14	BQ463121	BQ463121 HI04B23r
c 30	36.4	14.8	527	12	BF482462	BF482462 WHE1796_H
c 31	36.4	14.8	639	10	BE492050	BE492050 WHE0552_C
32	36.4	14.8	667	13	BI391466	BI391466 pgp1n.pk0
c 33	36.4	14.8	688	14	BQ801535	BQ801535 WHE2815_E
c 34	36.4	14.8	723	14	BQ281579	BQ281579 WHE3023_A
c 35	36.4	14.8	907	13	BI949635	BI949635 HVSME1001
c 36	36.4	14.8	914	17	CNS00CZP	AL059740 Drosophil
c 37	35.8	14.6	439	9	AL831137	AL831137 AL831137
38	35.8	14.6	667	17	AG045529	AG045529 Pan trogl
39	35.6	14.5	627	13	BM649599	BM649599 170006873
40	35.4	14.4	324	10	BE416124	BE416124 MUG004.H0
41	35.4	14.4	433	10	BE493608	BE493608 WHE0568_C
c 42	35.4	14.4	463	13	BJ305002	BJ305002 BJ305002
43	35.4	14.4	519	9	AL817756	AL817756 AL817756
44	35.4	14.4	530	10	BE498649	BE498649 WHE0964_D
45	35.4	14.4	637	14	BQ578407	BQ578407 WHE0302_E

ALIGNMENTS

RESULT 1
BM629517
LOCUS BM629517 668 bp mRNA linear EST 26-FEB-2002
DEFINITION 17000668640023 A.Gam.ad.cDNA1 Anopheles gambiae cDNA clone 19600448418736 5', mRNA sequence.
ACCESSION BM629517
VERSION BM629517.1 GI:18929028
KEYWORDS EST.
SOURCE African malaria mosquito.
ORGANISM Anopheles gambiae
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea; Anopheles.
REFERENCE 1 (bases 1 to 668)
AUTHORS Holt,R.A., Lin,J.-J., Murphy,S.D., Evans,C.A., Kraft,C.L., Charlab,R., Collins,F.H., Venter,J.C. and Hoffman,S.L.
TITLE Celera Anopheles gambiae EST project
JOURNAL Unpublished (2002)
COMMENT Contact: Holt R.A.
Celera Genomics
45 w. Gude Dr., Rockville, MD 20850, USA
Tel: 2404533151
Fax: 2404534580
Email: HoltRA@celera.com
Plate: NU010048I6 row: G column: 15
Seq primer: M13 Reverse.
FEATURES Location/Qualifiers
source 1. .668
/organism="Anopheles gambiae"

/strain="RSP-ST (Reduced susc. to Permethrin - std. chromosome)"
/db_xref="taxon:7165"
/clone="19600448418736"
/clone_lib="A.Gam.ad.cDNA1"
/dev_stage="Adult"
/lab_host="DH10b"
/note="Vector: pSport1; Site_1: Sal1; Site_2: Not1; Whole adult mosquitoes (mixed sex) frozen on liquid nitrogen. cDNA inserts >500 bp cloned directionally into pSport 1. Not 1 site is 3'. Clones available through the Malaria Research and Reference Reagent Resource Center (www.malaria.mr4.org)."

BASE COUNT 139 a 253 c 144 g 132 t
ORIGIN

Query Match 18.3%; Score 45; DB 13; Length 668;
Best Local Similarity 49.8%; Pred. No. 0.58;
Matches 114; Conservative 0; Mismatches 115; Indels 0; Gaps 0;

```
Qy          1 CTGGCCCATCAGACGTCGGCCGACCAACGGGCCGCATTGCTCGAGCTCGTACGAGACCAT 60
              | || |  | ||||   |||||  |||   ||||  |  |  |||   ||   |  | ||
Db        236 CCGGTCGTTAAGACTGTGGCCGTCCACTCGGCCCCGGTCGTCGCCGTCCACCACGCCCCG 295

Qy         61 GTGGCGGCAGTGCTCCGGCACGCGGACCCGAAAGCCATCGCGCCCGACCAGTCGTTCCGT 120
              || | |   |  |||| |  || | |||       |  |||  | | ||    || |||
Db        296 GTCGTGAAGGCCCTCCCGGTCGTGCACCACGCTCCGGTCGTCCACCACGCCCCGGTCCTG 355

Qy        121 GCACTCGGCTTCGATTCACTCACGGCCGTCGAGTTCCGAAACCTGCTGATCAAGGCAACA 180
                | | || | ||   | ||| |||| | | |    ||| | |      ||  |
Db        356 AAGACCGTCGTCCACTCTGCCCCGGTCGTCCACTCCGTGCACCCGGTCCCGGTGGTCAAG 415

Qy        181 GGACTCCGCCTTCCTGTCTCGCTGGTCTTCGACCACCCGACCCCTGCCA 229
                |  | |   || |   | || |  || || |||||   | || | |||
Db        416 AGCGTGCATGTTGCCAGCCCGGTCCTCGTCCACCACTAAAACCGTCCCA 464
```

RESULT 2
BM620935
LOCUS BM620935 672 bp mRNA linear EST 25-FEB-2002
DEFINITION 17000687445665 A.Gam.ad.cDNA1 Anopheles gambiae cDNA clone 19600449664641 5', mRNA sequence.
ACCESSION BM620935
VERSION BM620935.1 GI:18919353
KEYWORDS EST.
SOURCE African malaria mosquito.
ORGANISM Anopheles gambiae
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea; Anopheles.
REFERENCE 1 (bases 1 to 672)
AUTHORS Holt,R.A., Lin,J.-J., Murphy,S.D., Evans,C.A., Kraft,C.L., Charlab,R., Collins,F.H., Venter,J.C. and Hoffman,S.L.
TITLE Celera Anopheles gambiae EST project
JOURNAL Unpublished (2002)
COMMENT Contact: Holt R.A.
Celera Genomics
45 w. Gude Dr., Rockville, MD 20850, USA
Tel: 2404533151
Fax: 2404534580
Email: HoltRA@celera.com
Plate: NU01004I88 row: H column: 15
Seq primer: M13 Reverse.
FEATURES Location/Qualifiers
source 1. .672
/organism="Anopheles gambiae"
/strain="RSP-ST (Reduced susc. to Permethrin - std. chromosome)"
/db_xref="taxon:7165"
/clone="19600449664641"
/clone_lib="A.Gam.ad.cDNA1"
/dev_stage="Adult"
/lab_host="DH10b"
/note="Vector: pSport1; Site_1: Sal1; Site_2: Not1; Whole adult mosquitoes (mixed sex) frozen on liquid nitrogen. cDNA inserts >500 bp cloned directionally into pSport 1. Not 1 site is 3'. Clones available through the Malaria Research and Reference Reagent Resource Center (www.malaria.mr4.org)."

BASE COUNT 136 a 257 c 144 g 135 t
ORIGIN

Query Match 18.3%; Score 45; DB 13; Length 672;
Best Local Similarity 49.8%; Pred. No. 0.58;
Matches 114; Conservative 0; Mismatches 115; Indels 0; Gaps 0;

```
Qy          1 CTGGCCCATCAGACGTCGGCCGACCAACGGGCCGCATTGCTCGAGCTCGTACGAGACCAT 60
              | || |  | ||||   |||||  |||   ||||  |  |  |||   ||   |  | ||
Db        256 CCGGTCGTTAAGACTGTGGCCGTCCACTCGGCCCCGGTCGTCGCCGTCCACCACGCCCCG 315

Qy         61 GTGGCGGCAGTGCTCCGGCACGCGGACCCGAAAGCCATCGCGCCCGACCAGTCGTTCCGT 120
              || | |   |  |||| |  || | |||       |  |||  | | ||    || |||
Db        316 GTCGTGAAGGCCCTCCCGGTCGTGCACCACGCTCCGGTCGTCCACCACGCCCCGGTCCTG 375

Qy        121 GCACTCGGCTTCGATTCACTCACGGCCGTCGAGTTCCGAAACCTGCTGATCAAGGCAACA 180
                | | || | ||   | ||| |||| | | |    ||| | |      ||  |
Db        376 AAGACCGTCGTCCACTCTGCCCCGGTCGTCCACTCCGTGCACCCGGTCCCGGTGGTCAAG 435

Qy        181 GGACTCCGCCTTCCTGTCTCGCTGGTCTTCGACCACCCGACCCCTGCCA 229
                |  | |   || |   | || |  || || |||||   | || | |||
Db        436 AGCGTGCATGTTGCCAGCCCGGTCCTCGTCCACCACTAAAACCGTCCCA 484
```

RESULT 3
BM633427
LOCUS BM633427 687 bp mRNA linear EST 26-FEB-2002
DEFINITION 17000687507169 A.Gam.ad.cDNA1 Anopheles gambiae cDNA clone 19600449647458 5', mRNA sequence.
ACCESSION BM633427
VERSION BM633427.1 GI:18932938
KEYWORDS EST.
SOURCE African malaria mosquito.
ORGANISM Anopheles gambiae
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea; Anopheles.
REFERENCE 1 (bases 1 to 687)
AUTHORS Holt,R.A., Lin,J.-J., Murphy,S.D., Evans,C.A., Kraft,C.L., Charlab,R., Collins,F.H., Venter,J.C. and Hoffman,S.L.
TITLE Celera Anopheles gambiae EST project
JOURNAL Unpublished (2002)
COMMENT Contact: Holt R.A.
Celera Genomics
45 w. Gude Dr., Rockville, MD 20850, USA
Tel: 2404533151
Fax: 2404534580
Email: HoltRA@celera.com
Plate: NU010049W9 row: L column: 16
Seq primer: M13 Reverse.
FEATURES Location/Qualifiers
source 1. .687
/organism="Anopheles gambiae"
/strain="RSP-ST (Reduced susc. to Permethrin - std. chromosome)"
/db_xref="taxon:7165"
/clone="19600449647458"
/clone_lib="A.Gam.ad.cDNA1"
/dev_stage="Adult"
/lab_host="DH10b"
/note="Vector: pSport1; Site_1: Sal1; Site_2: Not1; Whole adult mosquitoes (mixed sex) frozen on liquid nitrogen. cDNA inserts >500 bp cloned directionally into pSport 1. Not 1 site is 3'. Clones available through the Malaria Research and Reference Reagent Resource Center (www.malaria.mr4.org)."
BASE COUNT 142 a 259 c 147 g 139 t

```
ORIGIN

  Query Match           18.3%;  Score 45;  DB 13;  Length 687;
  Best Local Similarity   49.8%;  Pred. No. 0.58;
  Matches  114;  Conservative    0;  Mismatches  115;  Indels    0;  Gaps    0;

Qy          1 CTGGCCCATCAGACGTCGGCCGACCAACGGGCCGCATTGCTCGAGCTCGTACGAGACCAT 60
              | || |  | ||||  ||||| |||  |||| |  |  |||  ||  |  | ||
Db        257 CCGGTCGTTAAGACTGTGGCCGTCCACTCGGCCCCGGTCGTCGCCGTCCACCACGCCCCG 316

Qy         61 GTGGCGGCAGTGCTCCGGCACGCGGACCCGAAAGCCATCGCGCCCGACCAGTCGTTCCGT 120
              || | |   |  |||| |  || | |||       |  |||  | | ||    || |||
Db        317 GTCGTGAAGGCCCTCCCGGTCGTGCACCACGCTCCGGTCGTCCACCACGCCCCGGTCCTG 376

Qy        121 GCACTCGGCTTCGATTCACTCACGGCCGTCGAGTTCCGAAACCTGCTGATCAAGGCAACA 180
              || | || | ||   | ||| |||| | | |    ||| | |       ||  |
Db        377 AAGACCGTCGTCCACTCTGCCCCGGTCGTCCACTCCGTGCACCCGGTCCCGGTGGTCAAG 436

Qy        181 GGACTCCGCCTTCCTGTCTCGCTGGTCTTCGACCACCCGACCCCTGCCA 229
                | | |   || |   | || |  || || |||||   | || | |||
Db        437 AGCGTGCATGTTGCCAGCCCGGTCCTCGTCCACCACTAAAACCGTCCCA 485


RESULT 4
BM650789
LOCUS       BM650789                698 bp    mRNA    linear   EST 26-FEB-2002
DEFINITION  17000687372551 A.Gam.ad.cDNA1 Anopheles gambiae cDNA clone
            19600449637497 5', mRNA sequence.
ACCESSION   BM650789
VERSION     BM650789.1  GI:18950300
KEYWORDS    EST.
SOURCE      African malaria mosquito.
  ORGANISM  Anopheles gambiae
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea;
            Anopheles.
REFERENCE   1  (bases 1 to 698)
  AUTHORS   Holt,R.A., Lin,J.-J., Murphy,S.D., Evans,C.A., Kraft,C.L., Charlab
            ,R., Collins,F.H., Venter,J.C. and Hoffman,S.L.
  TITLE     Celera Anopheles gambiae EST project
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Holt R.A.
            Celera Genomics
            45 w. Gude Dr., Rockville, MD 20850, USA
            Tel: 2404533151
            Fax: 2404534580
            Email: HoltRA@celera.com
            Plate: NU01004AYT  row: M  column: 15
            Seq primer: M13 Reverse.
FEATURES             Location/Qualifiers
     source          1. .698
                     /organism="Anopheles gambiae"
                     /strain="RSP-ST (Reduced susc. to Permethrin - std.
                     chromosome)"
                     /db_xref="taxon:7165"
                     /clone="19600449637497"
                     /clone_lib="A.Gam.ad.cDNA1"
                     /dev_stage="Adult"
                     /lab_host="DH10b"
                     /note="Vector: pSport1; Site_1: Sal1; Site_2: Not1; Whole
                     adult mosquitoes (mixed sex) frozen on liquid nitrogen.
                     cDNA  inserts >500 bp cloned directionally into pSport 1.
                     Not 1 site is  3'. Clones available through the Malaria
                     Research and Reference  Reagent Resource Center
                     (www.malaria.mr4.org)."
BASE COUNT      149 a    262 c    148 g    139 t
ORIGIN

  Query Match           18.3%;  Score 45;  DB 13;  Length 698;
  Best Local Similarity   49.8%;  Pred. No. 0.59;
  Matches  114;  Conservative    0;  Mismatches  115;  Indels    0;  Gaps    0;

Qy          1 CTGGCCCATCAGACGTCGGCCGACCAACGGGCCGCATTGCTCGAGCTCGTACGAGACCAT 60
              | || |  | ||||  ||||| |||  |||| |  |  |||  ||  |  | ||
Db        254 CCGGTCGTTAAGACTGTGGCCGTCCACTCGGCCCCGGTCGTCGCCGTCCACCACGCCCCG 313

Qy         61 GTGGCGGCAGTGCTCCGGCACGCGGACCCGAAAGCCATCGCGCCCGACCAGTCGTTCCGT 120
              || | |   |  |||| |  || | |||       |  |||  | | ||    || |||
Db        314 GTCGTGAAGGCCCTCCCGGTCGTGCACCACGCTCCGGTCGTCCACCACGCCCCGGTCCTG 373

Qy        121 GCACTCGGCTTCGATTCACTCACGGCCGTCGAGTTCCGAAACCTGCTGATCAAGGCAACA 180
              || | || | ||   | ||| |||| | | |    ||| | |       ||  |
Db        374 AAGACCGTCGTCCACTCTGCCCCGGTCGTCCACTCCGTGCACCCGGTCCCGGTGGTCAAG 433

Qy        181 GGACTCCGCCTTCCTGTCTCGCTGGTCTTCGACCACCCGACCCCTGCCA 229
                | | |   || |   | || |  || || |||||   | || | |||
Db        434 AGCGTGCATGTTGCCAGCCCGGTCCTCGTCCACCACTAAAACCGTCCCA 482


RESULT 5
BM653397
LOCUS       BM653397                715 bp    mRNA    linear   EST 26-FEB-2002
DEFINITION  17000687378407 A.Gam.ad.cDNA1 Anopheles gambiae cDNA clone
            19600449648803 5', mRNA sequence.
ACCESSION   BM653397
VERSION     BM653397.1  GI:18952908
KEYWORDS    EST.
SOURCE      African malaria mosquito.
  ORGANISM  Anopheles gambiae
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea;
            Anopheles.
REFERENCE   1  (bases 1 to 715)
  AUTHORS   Holt,R.A., Lin,J.-J., Murphy,S.D., Evans,C.A., Kraft,C.L., Charlab
            ,R., Collins,F.H., Venter,J.C. and Hoffman,S.L.
  TITLE     Celera Anopheles gambiae EST project
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Holt R.A.
            Celera Genomics
            45 w. Gude Dr., Rockville, MD 20850, USA
            Tel: 2404533151
            Fax: 2404534580
            Email: HoltRA@celera.com
            Plate: NU01004AYT  row: D  column: 17
            Seq primer: M13 Reverse.
FEATURES             Location/Qualifiers
     source          1. .715
                     /organism="Anopheles gambiae"
                     /strain="RSP-ST (Reduced susc. to Permethrin - std.
                     chromosome)"
                     /db_xref="taxon:7165"
                     /clone="19600449648803"
                     /clone_lib="A.Gam.ad.cDNA1"
                     /dev_stage="Adult"
                     /lab_host="DH10b"
                     /note="Vector: pSport1; Site_1: Sal1; Site_2: Not1; Whole
                     adult mosquitoes (mixed sex) frozen on liquid nitrogen.
                     cDNA  inserts >500 bp cloned directionally into pSport 1.
                     Not 1 site is  3'. Clones available through the Malaria
                     Research and Reference  Reagent Resource Center
                     (www.malaria.mr4.org)."
BASE COUNT      143 a    275 c    155 g    142 t
ORIGIN

  Query Match           18.3%;  Score 45;  DB 13;  Length 715;
  Best Local Similarity   49.8%;  Pred. No. 0.59;
  Matches  114;  Conservative    0;  Mismatches  115;  Indels    0;  Gaps    0;

Qy          1 CTGGCCCATCAGACGTCGGCCGACCAACGGGCCGCATTGCTCGAGCTCGTACGAGACCAT 60
              | || |  | ||||  ||||| |||  |||| |  |  |||  ||  |  | ||
Db        285 CCGGTCGTTAAGACTGTGGCCGTCCACTCGGCCCCGGTCGTCGCCGTCCACCACGCCCCG 344

Qy         61 GTGGCGGCAGTGCTCCGGCACGCGGACCCGAAAGCCATCGCGCCCGACCAGTCGTTCCGT 120
              || | |   |  |||| |  || | |||       |  |||  | | ||    || |||
Db        345 GTCGTGAAGGCCCTCCCGGTCGTGCACCACGCTCCGGTCGTCCACCACGCCCCGGTCCTG 404
```

```
Qy        121 GCACTCGGCTTCGATTCACTCACGGCCGTCGAGTTCCGAAACCTGCTGATCAAGGCAACA 180
              || | || | ||   | ||| |||| | | |    ||| | |      ||  |
Db        405 AAGACCGTCGTCCACTCTGCCCCGGTCGTCCACTCCGTGCACCCGGTCCCGGTGGTCAAG 464

Qy        181 GGACTCCGCCTTCCTGTCTCGCTGGTCTTCGACCACCCGACCCCTGCCA 229
                |  | |   || |  | || |  || || |||||   | || | |||
Db        465 AGCGTGCATGTTGCCAGCCCGGTCCTCGTCCACCACTAAAACCGTCCCA 513


RESULT 6
BM638546
LOCUS       BM638546                509 bp    mRNA    linear   EST 26-FEB-2002
DEFINITION  17000687566790 A.Gam.ad.cDNA1 Anopheles gambiae cDNA clone
            19600449667064 5', mRNA sequence.
ACCESSION   BM638546
VERSION     BM638546.1  GI:18938057
KEYWORDS    EST.
SOURCE      African malaria mosquito.
  ORGANISM  Anopheles gambiae
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea;
            Anopheles.
REFERENCE   1  (bases 1 to 509)
  AUTHORS   Holt,R.A., Lin,J.-J., Murphy,S.D., Evans,C.A., Kraft,C.L., Charlab
            ,R., Collins,F.H., Venter,J.C. and Hoffman,S.L.
  TITLE     Celera Anopheles gambiae EST project
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Holt R.A.
            Celera Genomics
            45 w. Gude Dr., Rockville, MD 20850, USA
            Tel: 2404533151
            Fax: 2404534580
            Email: HoltRA@celera.com
            Plate: NU01004HMK  row: M  column: 14
            Seq primer: M13 Reverse.
FEATURES             Location/Qualifiers
     source          1. .509
                     /organism="Anopheles gambiae"
                     /strain="RSP-ST (Reduced susc. to Permethrin - std.
                     chromosome)"
                     /db_xref="taxon:7165"
                     /clone="19600449667064"
                     /clone_lib="A.Gam.ad.cDNA1"
                     /dev_stage="Adult"
                     /lab_host="DH10b"
                     /note="Vector: pSport1; Site_1: Sal1; Site_2: Not1; Whole
                     adult mosquitoes (mixed sex) frozen on liquid nitrogen.
                     cDNA  inserts >500 bp cloned directionally into pSport 1.
                     Not 1 site is  3'. Clones available through the Malaria
                     Research and Reference  Reagent Resource Center
                     (www.malaria.mr4.org)."
BASE COUNT       81 a    207 c    116 g    105 t
ORIGIN

  Query Match           17.6%;  Score 43.4;  DB 13;  Length 509;
  Best Local Similarity   49.3%;  Pred. No. 1.3;
  Matches  113;  Conservative    0;  Mismatches  116;  Indels    0;  Gaps    0;

Qy          1 CTGGCCCATCAGACGTCGGCCGACCAACGGGCCGCATTGCTCGAGCTCGTACGAGACCAT 60
              | || |  | ||||    |||| |||   |||| |  |  |||   ||    |  | ||
Db        259 CCGGTCGTTAAGACTGTTGCCGTCCACTCGGCCCCGGTCGTCGCCGTCCACCACGCCCCG 318

Qy         61 GTGGCGGCAGTGCTCCGGCACGCGGACCCGAAAGCCATCGCGCCCGACCAGTCGTTCCGT 120
              || | |   |  |||| |  || | |||       |  |||  | | ||    || |||
Db        319 GTCGTGAAGGCCCTCCCGGTCGTGCACCACGCTCCGGTCGTCCACCACGCCCCGGTCCTG 378

Qy        121 GCACTCGGCTTCGATTCACTCACGGCCGTCGAGTTCCGAAACCTGCTGATCAAGGCAACA 180
              || | || | ||   | ||| |||| | | |    ||| | |      ||  |
Db        379 AAGACCGTCGTCCACTCTGCCCCGGTCGTCCACTCCGTGCACCCGGTCCCGGTGGTCAAG 438

Qy        181 GGACTCCGCCTTCCTGTCTCGCTGGTCTTCGACCACCCGACCCCTGCCA 229
                |  | |   || |  | || |  || || |||||   | || | |||
Db        439 AGCGTGCATGTTGCCAGCCCGGTCCTCGTCCACCACTAAAACCGTCCCA 487


RESULT 7
BM609994
LOCUS       BM609994                549 bp    mRNA    linear   EST 25-FEB-2002
DEFINITION  17000687108219 A.Gam.ad.cDNA.blood1 Anopheles gambiae cDNA clone
            19600449680139 5', mRNA sequence.
ACCESSION   BM609994
VERSION     BM609994.1  GI:18908086
KEYWORDS    EST.
SOURCE      African malaria mosquito.
  ORGANISM  Anopheles gambiae
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea;
            Anopheles.
REFERENCE   1  (bases 1 to 549)
  AUTHORS   Holt,R.A., Lin,J.-J., Murphy,S.D., Evans,C.A., Kraft,C.L., Charlab
            ,R., Collins,F.H., Venter,J.C. and Hoffman,S.L.
  TITLE     Celera Anopheles gambiae EST project
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Holt R.A.
            Celera Genomics
            45 w. Gude Dr., Rockville, MD 20850, USA
            Tel: 2404533151
            Fax: 2404534580
            Email: HoltRA@celera.com
            Plate: NU01004B3L  row: M  column: 21
            Seq primer: M13 Reverse.
FEATURES             Location/Qualifiers
     source          1. .549
                     /organism="Anopheles gambiae"
                     /strain="RSP-ST (Reduced susc. to Permethrin - std.
                     chromosome)"
                     /db_xref="taxon:7165"
                     /clone="19600449680139"
                     /clone_lib="A.Gam.ad.cDNA.blood1"
                     /dev_stage="Adult"
                     /lab_host="DH10b"
                     /note="Vector: pSport1; Site_1: Sal1; Site_2: Not1; Whole
                     adult mosquitoes (mixed sex) frozen on liquid nitrogen 24
                     hours  after human blood feeding. cDNA inserts >500 bp
                     cloned directionally  into pSport 1. Not 1 site is 3'.
                     Clones available through the Malaria  Research and
                     Reference Reagent Resource Center (www.malaria.mr4.org)"
BASE COUNT       97 a    216 c    122 g    114 t
ORIGIN

  Query Match           17.6%;  Score 43.4;  DB 13;  Length 549;
  Best Local Similarity   49.3%;  Pred. No. 1.3;
  Matches  113;  Conservative    0;  Mismatches  116;  Indels    0;  Gaps    0;

Qy          1 CTGGCCCATCAGACGTCGGCCGACCAACGGGCCGCATTGCTCGAGCTCGTACGAGACCAT 60
              | || |  | ||||    |||| |||   |||| |  |  |||   ||    |  | ||
Db        253 CCGGTCGTTAAGACTGTTGCCGTCCACTCGGCCCCGGTCGTCGCCGTCCACCACGCCCCG 312

Qy         61 GTGGCGGCAGTGCTCCGGCACGCGGACCCGAAAGCCATCGCGCCCGACCAGTCGTTCCGT 120
              || | |   |  |||| |  || | |||       |  |||  | | ||    || |||
Db        313 GTCGTGAAGGCCCTCCCGGTCGTGCACCACGCTCCGGTCGTCCACCACGCCCCGGTCCTG 372

Qy        121 GCACTCGGCTTCGATTCACTCACGGCCGTCGAGTTCCGAAACCTGCTGATCAAGGCAACA 180
              || | || | ||   | ||| |||| | | |    ||| | |      ||  |
Db        373 AAGACCGTCGTCCACTCTGCCCCGGTCGTCCACTCCGTGCACCCGGTCCCGGTGGTCAAG 432

Qy        181 GGACTCCGCCTTCCTGTCTCGCTGGTCTTCGACCACCCGACCCCTGCCA 229
                |  | |   || |  | || |  || || |||||   | || | |||
Db        433 AGCGTGCATGTTGCCAGCCCGGTCCTCGTCCACCACTAAAACCGTCCCA 481


RESULT 8
BM649092
LOCUS       BM649092                567 bp    mRNA    linear   EST 26-FEB-2002
DEFINITION  17000687368208 A.Gam.ad.cDNA1 Anopheles gambiae cDNA clone
            19600449667646 5', mRNA sequence.
```

```
ACCESSION   BM649092
VERSION     BM649092.1  GI:18948603
KEYWORDS    EST.
SOURCE      African malaria mosquito.
  ORGANISM  Anopheles gambiae
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea;
            Anopheles.
REFERENCE   1  (bases 1 to 567)
  AUTHORS   Holt,R.A., Lin,J.-J., Murphy,S.D., Evans,C.A., Kraft,C.L., Charlab
            ,R., Collins,F.H., Venter,J.C. and Hoffman,S.L.
  TITLE     Celera Anopheles gambiae EST project
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Holt R.A.
            Celera Genomics
            45 w. Gude Dr., Rockville, MD 20850, USA
            Tel: 2404533151
            Fax: 2404534580
            Email: HoltRA@celera.com
            Plate: NU010049UF  row: E  column: 20
            Seq primer: M13 Reverse.
FEATURES             Location/Qualifiers
     source          1. .567
                     /organism="Anopheles gambiae"
                     /strain="RSP-ST (Reduced susc. to Permethrin - std.
                     chromosome)"
                     /db_xref="taxon:7165"
                     /clone="19600449667646"
                     /clone_lib="A.Gam.ad.cDNA1"
                     /dev_stage="Adult"
                     /lab_host="DH10b"
                     /note="Vector: pSport1; Site_1: Sal1; Site_2: Not1; Whole
                     adult mosquitoes (mixed sex) frozen on liquid nitrogen.
                     cDNA  inserts >500 bp cloned directionally into pSport 1.
                     Not 1 site is  3'. Clones available through the Malaria
                     Research and Reference  Reagent Resource Center
                     (www.malaria.mr4.org)."
BASE COUNT       98 a    220 c    130 g    119 t
ORIGIN

  Query Match           17.6%;  Score 43.4;  DB 13;  Length 567;
  Best Local Similarity   49.3%;  Pred. No. 1.3;
  Matches  113;  Conservative    0;  Mismatches  116;  Indels    0;  Gaps    0;

Qy          1 CTGGCCCATCAGACGTCGGCCGACCAACGGGCCGCATTGCTCGAGCTCGTACGAGACCAT 60
              | || |   | ||||    |||| |||   ||||  |   |  |||   ||   |  | ||
Db        278 CCGGTCGTTAAGACTGTTGCCGTCCACTCGGCCCCGGTCGTCGCCGTCCACCACGCCCCG 337

Qy         61 GTGGCGGCAGTGCTCCGGCACGCGGACCCGAAAGCCATCGCGCCCGACCAGTCGTTCCGT 120
              || | |   |  |||| |  || | |||     | |||  | | ||    || |||
Db        338 GTCGTGAAGGCCCTCCCGGTCGTGCACCACGCTCCGGTCGTCCACCACGCCCCGGTCCTG 397

Qy        121 GCACTCGGCTTCGATTCACTCACGGCCGTCGAGTTCCGAAACCTGCTGATCAAGGCAACA 180
              || | ||  | ||   | |||  |||| | | |    |||  | |       ||  |
Db        398 AAGACCGTCGTCCACTCTGCCCCGGTCGTCCACTCCGTGCACCCGGTCCCGGTGGTCAAG 457

Qy        181 GGACTCCGCCTTCCTGTCTCGCTGGTCTTCGACCACCCGACCCCTGCCA 229
                |  | |   || |   | || |  || || |||||   | || | |||
Db        458 AGCGTGCATGTTGCCAGCCCGGTCCTCGTCCACCACTAAAACCGTCCCA 506


RESULT 9
BM632295
LOCUS       BM632295                602 bp    mRNA    linear   EST 26-FEB-2002
DEFINITION  17000687504066 A.Gam.ad.cDNA1 Anopheles gambiae cDNA clone
            19600449668028 5', mRNA sequence.
ACCESSION   BM632295
VERSION     BM632295.1  GI:18931806
KEYWORDS    EST.
SOURCE      African malaria mosquito.
  ORGANISM  Anopheles gambiae
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea;
            Anopheles.
REFERENCE   1  (bases 1 to 602)
  AUTHORS   Holt,R.A., Lin,J.-J., Murphy,S.D., Evans,C.A., Kraft,C.L., Charlab
            ,R., Collins,F.H., Venter,J.C. and Hoffman,S.L.
  TITLE     Celera Anopheles gambiae EST project
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Holt R.A.
            Celera Genomics
            45 w. Gude Dr., Rockville, MD 20850, USA
            Tel: 2404533151
            Fax: 2404534580
            Email: HoltRA@celera.com
            Plate: NU01004ABH  row: E  column: 18
            Seq primer: M13 Reverse.
FEATURES             Location/Qualifiers
     source          1. .602
                     /organism="Anopheles gambiae"
                     /strain="RSP-ST (Reduced susc. to Permethrin - std.
                     chromosome)"
                     /db_xref="taxon:7165"
                     /clone="19600449668028"
                     /clone_lib="A.Gam.ad.cDNA1"
                     /dev_stage="Adult"
                     /lab_host="DH10b"
                     /note="Vector: pSport1; Site_1: Sal1; Site_2: Not1; Whole
                     adult mosquitoes (mixed sex) frozen on liquid nitrogen.
                     cDNA  inserts >500 bp cloned directionally into pSport 1.
                     Not 1 site is  3'. Clones available through the Malaria
                     Research and Reference  Reagent Resource Center
                     (www.malaria.mr4.org)."
BASE COUNT      115 a    240 c    136 g    111 t
ORIGIN

  Query Match           17.6%;  Score 43.4;  DB 13;  Length 602;
  Best Local Similarity   49.3%;  Pred. No. 1.4;
  Matches  113;  Conservative    0;  Mismatches  116;  Indels    0;  Gaps    0;

Qy          1 CTGGCCCATCAGACGTCGGCCGACCAACGGGCCGCATTGCTCGAGCTCGTACGAGACCAT 60
              | || |   | ||||    |||| |||   ||||  |   |  |||   ||   |  | ||
Db        160 CCGGTCGTTAAGACTGTTGCCGTCCACTCGGCCCCGGTCGTCGCCGTCCACCACGCCCCG 219

Qy         61 GTGGCGGCAGTGCTCCGGCACGCGGACCCGAAAGCCATCGCGCCCGACCAGTCGTTCCGT 120
              || | |   |  |||| |  || | |||     | |||  | | ||    || |||
Db        220 GTCGTGAAGGCCCTCCCGGTCGTGCACCACGCTCCGGTCGTCCACCACGCCCCGGTCCTG 279

Qy        121 GCACTCGGCTTCGATTCACTCACGGCCGTCGAGTTCCGAAACCTGCTGATCAAGGCAACA 180
              || | ||  | ||   | |||  |||| | | |    |||  | |       ||  |
Db        280 AAGACCGTCGTCCACTCTGCCCCGGTCGTCCACTCCGTGCACCCGGTCCCGGTGGTCAAG 339

Qy        181 GGACTCCGCCTTCCTGTCTCGCTGGTCTTCGACCACCCGACCCCTGCCA 229
                |  | |   || |   | || |  || || |||||   | || | |||
Db        340 AGCGTGCATGTTGCCAGCCCGGTCCTCGTCCACCACTAAAACCGTCCCA 388


RESULT 10
BM632995
LOCUS       BM632995                639 bp    mRNA    linear   EST 26-FEB-2002
DEFINITION  17000687505513 A.Gam.ad.cDNA1 Anopheles gambiae cDNA clone
            19600449633160 5', mRNA sequence.
ACCESSION   BM632995
VERSION     BM632995.1  GI:18932506
KEYWORDS    EST.
SOURCE      African malaria mosquito.
  ORGANISM  Anopheles gambiae
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea;
            Anopheles.
REFERENCE   1  (bases 1 to 639)
  AUTHORS   Holt,R.A., Lin,J.-J., Murphy,S.D., Evans,C.A., Kraft,C.L., Charlab
            ,R., Collins,F.H., Venter,J.C. and Hoffman,S.L.
  TITLE     Celera Anopheles gambiae EST project
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Holt R.A.
```

```
            Celera Genomics
            45 w. Gude Dr., Rockville, MD 20850, USA
            Tel: 2404533151
            Fax: 2404534580
            Email: HoltRA@celera.com
            Plate: NU01004ABY  row: H  column: 22
            Seq primer: M13 Reverse.
FEATURES              Location/Qualifiers
     source           1. .639
                      /organism="Anopheles gambiae"
                      /strain="RSP-ST (Reduced susc. to Permethrin - std.
                      chromosome)"
                      /db_xref="taxon:7165"
                      /clone="19600449633160"
                      /clone_lib="A.Gam.ad.cDNA1"
                      /dev_stage="Adult"
                      /lab_host="DH10b"
                      /note="Vector: pSport1; Site_1: Sal1; Site_2: Not1; Whole
                      adult mosquitoes (mixed sex) frozen on liquid nitrogen.
                      cDNA  inserts >500 bp cloned directionally into pSport 1.
                      Not 1 site is  3'. Clones available through the Malaria
                      Research and Reference  Reagent Resource Center
                      (www.malaria.mr4.org)."
BASE COUNT      119 a    251 c    140 g    129 t
ORIGIN

  Query Match             17.6%;  Score 43.4;  DB 13;  Length 639;
  Best Local Similarity   49.3%;  Pred. No. 1.4;
  Matches  113;  Conservative    0;  Mismatches  116;  Indels    0;  Gaps    0;

Qy          1 CTGGCCCATCAGACGTCGGCCGACCAACGGGCCGCATTGCTCGAGCTCGTACGAGACCAT 60
              | || |  | ||||    |||| |||   |||| |  |  |||   ||   |  | ||
Db        262 CCGGTCGTTAAGACTGTTGCCGTCCACTCGGCCCCGGTCGTCGCCGTCCACCACGCCCCG 321

Qy         61 GTGGCGGCAGTGCTCCGGCACGCGGACCCGAAAGCCATCGCGCCCGACCAGTCGTTCCGT 120
              || | |   |  |||| |  || | |||      |  |||  | | ||    || |||
Db        322 GTCGTGAAGGCCCTCCCGGTCGTGCACCACGCTCCGGTCGTCCACCACGCCCCGGTCCTG 381

Qy        121 GCACTCGGCTTCGATTCACTCACGGCCGTCGAGTTCCGAAACCTGCTGATCAAGGCAACA 180
                || | || | ||   | ||| |||| | | |    ||| | |      ||  |
Db        382 AAGACCGTCGTCCACTCTGCCCCGGTCGTCCACTCCGTGCACCCGGTCCCGGTGGTCAAG 441

Qy        181 GGACTCCGCCTTCCTGTCTCGCTGGTCTTCGACCACCCGACCCCTGCCA 229
              |  | |   || |   | || |  || || |||||   | || | |||
Db        442 AGCGTGCATGTTGCCAGCCCGGTCCTCGTCCACCACTAAAACCGTCCCA 490


RESULT 11
BM645157
LOCUS       BM645157                663 bp    mRNA    linear   EST 26-FEB-2002
DEFINITION  17000687315754 A.Gam.ad.cDNA1 Anopheles gambiae cDNA clone
            19600449664221 5', mRNA sequence.
ACCESSION   BM645157
VERSION     BM645157.1  GI:18944668
KEYWORDS    EST.
SOURCE      African malaria mosquito.
  ORGANISM  Anopheles gambiae
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea;
            Anopheles.
REFERENCE   1  (bases 1 to 663)
  AUTHORS   Holt,R.A., Lin,J.-J., Murphy,S.D., Evans,C.A., Kraft,C.L., Charlab
            ,R., Collins,F.H., Venter,J.C. and Hoffman,S.L.
  TITLE     Celera Anopheles gambiae EST project
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Holt R.A.
            Celera Genomics
            45 w. Gude Dr., Rockville, MD 20850, USA
            Tel: 2404533151
            Fax: 2404534580
            Email: HoltRA@celera.com
            Plate: NU01004I7U  row: G  column: 03
            Seq primer: M13 Reverse.
FEATURES              Location/Qualifiers
     source           1. .663
                      /organism="Anopheles gambiae"
                      /strain="RSP-ST (Reduced susc. to Permethrin - std.
                      chromosome)"
                      /db_xref="taxon:7165"
                      /clone="19600449664221"
                      /clone_lib="A.Gam.ad.cDNA1"
                      /dev_stage="Adult"
                      /lab_host="DH10b"
                      /note="Vector: pSport1; Site_1: Sal1; Site_2: Not1; Whole
                      adult mosquitoes (mixed sex) frozen on liquid nitrogen.
                      cDNA  inserts >500 bp cloned directionally into pSport 1.
                      Not 1 site is  3'. Clones available through the Malaria
                      Research and Reference  Reagent Resource Center
                      (www.malaria.mr4.org)."
BASE COUNT      136 a    254 c    142 g    131 t
ORIGIN

  Query Match             17.6%;  Score 43.4;  DB 13;  Length 663;
  Best Local Similarity   49.3%;  Pred. No. 1.4;
  Matches  113;  Conservative    0;  Mismatches  116;  Indels    0;  Gaps    0;

Qy          1 CTGGCCCATCAGACGTCGGCCGACCAACGGGCCGCATTGCTCGAGCTCGTACGAGACCAT 60
              | || |  | ||||    |||| |||   |||| |  |  |||   ||   |  | ||
Db        236 CCGGTCGTTAAGACTGTTGCCGTCCACTCGGCCCCGGTCGTCGCCGTCCACCACGCCCCG 295

Qy         61 GTGGCGGCAGTGCTCCGGCACGCGGACCCGAAAGCCATCGCGCCCGACCAGTCGTTCCGT 120
              || | |   |  |||| |  || | |||      |  |||  | | ||    || |||
Db        296 GTCGTGAAGGCCCTCCCGGTCGTGCACCACGCTCCGGTCGTCCACCACGCCCCGGTCCTG 355

Qy        121 GCACTCGGCTTCGATTCACTCACGGCCGTCGAGTTCCGAAACCTGCTGATCAAGGCAACA 180
                || | || | ||   | ||| |||| | | |    ||| | |      ||  |
Db        356 AAGACCGTCGTCCACTCTGCCCCGGTCGTCCACTCCGTGCACCCGGTCCCGGTGGTCAAG 415

Qy        181 GGACTCCGCCTTCCTGTCTCGCTGGTCTTCGACCACCCGACCCCTGCCA 229
              |  | |   || |   | || |  || || |||||   | || | |||
Db        416 AGCGTGCATGTTGCCAGCCCGGTCCTCGTCCACCACTAAAACCGTCCCA 464


RESULT 12
BM627481
LOCUS       BM627481                678 bp    mRNA    linear   EST 26-FEB-2002
DEFINITION  17000687496976 A.Gam.ad.cDNA1 Anopheles gambiae cDNA clone
            19600449639797 5', mRNA sequence.
ACCESSION   BM627481
VERSION     BM627481.1  GI:18926992
KEYWORDS    EST.
SOURCE      African malaria mosquito.
  ORGANISM  Anopheles gambiae
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea;
            Anopheles.
REFERENCE   1  (bases 1 to 678)
  AUTHORS   Holt,R.A., Lin,J.-J., Murphy,S.D., Evans,C.A., Kraft,C.L., Charlab
            ,R., Collins,F.H., Venter,J.C. and Hoffman,S.L.
  TITLE     Celera Anopheles gambiae EST project
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Holt R.A.
            Celera Genomics
            45 w. Gude Dr., Rockville, MD 20850, USA
            Tel: 2404533151
            Fax: 2404534580
            Email: HoltRA@celera.com
            Plate: NU01004AYA  row: M  column: 11
            Seq primer: M13 Reverse.
FEATURES              Location/Qualifiers
     source           1. .678
                      /organism="Anopheles gambiae"
                      /strain="RSP-ST (Reduced susc. to Permethrin - std.
                      chromosome)"
                      /db_xref="taxon:7165"
                      /clone="19600449639797"
```

```
                     /clone_lib="A.Gam.ad.cDNA1"
                     /dev_stage="Adult"
                     /lab_host="DH10b"
                     /note="Vector: pSport1; Site_1: Sal1; Site_2: Not1; Whole
                     adult mosquitoes (mixed sex) frozen on liquid nitrogen.
                     cDNA  inserts >500 bp cloned directionally into pSport 1.
                     Not 1 site is  3'. Clones available through the Malaria
                     Research and Reference  Reagent Resource Center
                     (www.malaria.mr4.org)."
BASE COUNT      139 a    258 c    145 g    136 t
ORIGIN

  Query Match           17.6%;  Score 43.4;  DB 13;  Length 678;
  Best Local Similarity   49.3%;  Pred. No. 1.4;
  Matches  113;  Conservative    0;  Mismatches  116;  Indels    0;  Gaps    0;

Qy          1 CTGGCCCATCAGACGTCGGCCGACCAACGGGCCGCATTGCTCGAGCTCGTACGAGACCAT 60
              | || |  | ||||    |||| |||   |||| |  |  |||   ||   |  | ||
Db        251 CCGGTCGTTAAGACTGTTGCCGTCCACTCGGCCCCGGTCGTCGCCGTCCACCACGCCCCG 310

Qy         61 GTGGCGGCAGTGCTCCGGCACGCGGACCCGAAAGCCATCGCGCCCGACCAGTCGTTCCGT 120
              || | |   |  |||| |  || | |||       |  |||  | | ||    || |||
Db        311 GTCGTGAAGGCCCTCCCGGTCGTGCACCACGCTCCGGTCGTCCACCACGCCCCGGTCCTG 370

Qy        121 GCACTCGGCTTCGATTCACTCACGGCCGTCGAGTTCCGAAACCTGCTGATCAAGGCAACA 180
              || | || | ||   | ||| |||| | | |    |||  | |     ||  |
Db        371 AAGACCGTCGTCCACTCTGCCCCGGTCGTCCACTCCGTGCACCCGGTCCCGGTGGTCAAG 430

Qy        181 GGACTCCGCCTTCCTGTCTCGCTGGTCTTCGACCACCCGACCCCTGCCA 229
              |  | |   || |   | || |  || || |||||   | || | |||
Db        431 AGCGTGCATGTTGCCAGCCCGGTCCTCGTCCACCACTAAAACCGTCCCA 479


RESULT 13
BM656823
LOCUS       BM656823                678 bp    mRNA    linear   EST 26-FEB-2002
DEFINITION  17000687389299 A.Gam.ad.cDNA1 Anopheles gambiae cDNA clone
            19600449659034 5', mRNA sequence.
ACCESSION   BM656823
VERSION     BM656823.1  GI:18956334
KEYWORDS    EST.
SOURCE      African malaria mosquito.
  ORGANISM  Anopheles gambiae
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea;
            Anopheles.
REFERENCE   1  (bases 1 to 678)
  AUTHORS   Holt,R.A., Lin,J.-J., Murphy,S.D., Evans,C.A., Kraft,C.L., Charlab
            ,R., Collins,F.H., Venter,J.C. and Hoffman,S.L.
  TITLE     Celera Anopheles gambiae EST project
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Holt R.A.
            Celera Genomics
            45 w. Gude Dr., Rockville, MD 20850, USA
            Tel: 2404533151
            Fax: 2404534580
            Email: HoltRA@celera.com
            Plate: NU01004HNI  row: N  column: 24
            Seq primer: M13 Reverse.
FEATURES             Location/Qualifiers
     source          1. .678
                     /organism="Anopheles gambiae"
                     /strain="RSP-ST (Reduced susc. to Permethrin - std.
                     chromosome)"
                     /db_xref="taxon:7165"
                     /clone="19600449659034"
                     /clone_lib="A.Gam.ad.cDNA1"
                     /dev_stage="Adult"
                     /lab_host="DH10b"
                     /note="Vector: pSport1; Site_1: Sal1; Site_2: Not1; Whole
                     adult mosquitoes (mixed sex) frozen on liquid nitrogen.
                     cDNA  inserts >500 bp cloned directionally into pSport 1.
                     Not 1 site is  3'. Clones available through the Malaria
                     Research and Reference  Reagent Resource Center
                     (www.malaria.mr4.org)."
BASE COUNT      137 a    259 c    146 g    136 t
ORIGIN

  Query Match           17.6%;  Score 43.4;  DB 13;  Length 678;
  Best Local Similarity   49.3%;  Pred. No. 1.4;
  Matches  113;  Conservative    0;  Mismatches  116;  Indels    0;  Gaps    0;

Qy          1 CTGGCCCATCAGACGTCGGCCGACCAACGGGCCGCATTGCTCGAGCTCGTACGAGACCAT 60
              | || |  | ||||    |||| |||   |||| |  |  |||   ||   |  | ||
Db        260 CCGGTCGTTAAGACTGTTGCCGTCCACTCGGCCCCGGTCGTCGCCGTCCACCACGCCCCG 319

Qy         61 GTGGCGGCAGTGCTCCGGCACGCGGACCCGAAAGCCATCGCGCCCGACCAGTCGTTCCGT 120
              || | |   |  |||| |  || | |||       |  |||  | | ||    || |||
Db        320 GTCGTGAAGGCCCTCCCGGTCGTGCACCACGCTCCGGTCGTCCACCACGCCCCGGTCCTG 379

Qy        121 GCACTCGGCTTCGATTCACTCACGGCCGTCGAGTTCCGAAACCTGCTGATCAAGGCAACA 180
              || | || | ||   | ||| |||| | | |    |||  | |     ||  |
Db        380 AAGACCGTCGTCCACTCTGCCCCGGTCGTCCACTCCGTGCACCCGGTCCCGGTGGTCAAG 439

Qy        181 GGACTCCGCCTTCCTGTCTCGCTGGTCTTCGACCACCCGACCCCTGCCA 229
              |  | |   || |   | || |  || || |||||   | || | |||
Db        440 AGCGTGCATGTTGCCAGCCCGGTCCTCGTCCACCACTAAAACCGTCCCA 488


RESULT 14
BM634614
LOCUS       BM634614                684 bp    mRNA    linear   EST 26-FEB-2002
DEFINITION  17000687554700 A.Gam.ad.cDNA1 Anopheles gambiae cDNA clone
            19600449655338 5', mRNA sequence.
ACCESSION   BM634614
VERSION     BM634614.1  GI:18934125
KEYWORDS    EST.
SOURCE      African malaria mosquito.
  ORGANISM  Anopheles gambiae
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea;
            Anopheles.
REFERENCE   1  (bases 1 to 684)
  AUTHORS   Holt,R.A., Lin,J.-J., Murphy,S.D., Evans,C.A., Kraft,C.L., Charlab
            ,R., Collins,F.H., Venter,J.C. and Hoffman,S.L.
  TITLE     Celera Anopheles gambiae EST project
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Holt R.A.
            Celera Genomics
            45 w. Gude Dr., Rockville, MD 20850, USA
            Tel: 2404533151
            Fax: 2404534580
            Email: HoltRA@celera.com
            Plate: NU01004A9Q  row: D  column: 24
            Seq primer: M13 Reverse.
FEATURES             Location/Qualifiers
     source          1. .684
                     /organism="Anopheles gambiae"
                     /strain="RSP-ST (Reduced susc. to Permethrin - std.
                     chromosome)"
                     /db_xref="taxon:7165"
                     /clone="19600449655338"
                     /clone_lib="A.Gam.ad.cDNA1"
                     /dev_stage="Adult"
                     /lab_host="DH10b"
                     /note="Vector: pSport1; Site_1: Sal1; Site_2: Not1; Whole
                     adult mosquitoes (mixed sex) frozen on liquid nitrogen.
                     cDNA  inserts >500 bp cloned directionally into pSport 1.
                     Not 1 site is  3'. Clones available through the Malaria
                     Research and Reference  Reagent Resource Center
                     (www.malaria.mr4.org)."
BASE COUNT      139 a    260 c    146 g    139 t
ORIGIN

  Query Match           17.6%;  Score 43.4;  DB 13;  Length 684;
  Best Local Similarity   49.3%;  Pred. No. 1.4;
```

```
  Matches  113;  Conservative    0;  Mismatches  116;  Indels    0;  Gaps    0;

Qy          1 CTGGCCCATCAGACGTCGGCCGACCAACGGGCCGCATTGCTCGAGCTCGTACGAGACCAT 60
              | || |  | ||||    |||| |||   |||| |  |  |||   ||   |  | ||
Db        258 CCGGTCGTTAAGACTGTTGCCGTCCACTCGGCCCCGGTCGTCGCCGTCCACCACGCCCCG 317

Qy         61 GTGGCGGCAGTGCTCCGGCACGCGGACCCGAAAGCCATCGCGCCCGACCAGTCGTTCCGT 120
              || |  |   |||| |  || | |||       | |||  | | ||    || |||
Db        318 GTCGTGAAGGCCCTCCCGGTCGTGCACCACGCTCCGGTCGTCCACCACGCCCCGGTCCTG 377

Qy        121 GCACTCGGCTTCGATTCACTCACGGCCGTCGAGTTCCGAAACCTGCTGATCAAGGCAACA 180
              || | || | ||   | ||| |||| | | |   ||| | |      || |
Db        378 AAGACCGTCGTCCACTCTGCCCCGGTCGTCCACTCCGTGCACCCGGTCCCGGTGGTCAAG 437

Qy        181 GGACTCCGCCTTCCTGTCTCGCTGGTCTTCGACCACCCGACCCCTGCCA 229
                |  | |   || |   | || |  || || |||||   | || | |||
Db        438 AGCGTGCATGTTGCCAGCCCGGTCCTCGTCCACCACTAAAACCGTCCCA 486


RESULT 15
BM648344
LOCUS       BM648344                684 bp    mRNA    linear   EST 26-FEB-2002
DEFINITION  17000687325792 A.Gam.ad.cDNA1 Anopheles gambiae cDNA clone
            19600449628524 5', mRNA sequence.
ACCESSION   BM648344
VERSION     BM648344.1  GI:18947867
KEYWORDS    EST.
SOURCE      African malaria mosquito.
  ORGANISM  Anopheles gambiae
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea;
            Anopheles.
REFERENCE   1  (bases 1 to 684)
  AUTHORS   Holt,R.A., Lin,J.-J., Murphy,S.D., Evans,C.A., Kraft,C.L., Charlab
            ,R., Collins,F.H., Venter,J.C. and Hoffman,S.L.
  TITLE     Celera Anopheles gambiae EST project
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Holt R.A.
            Celera Genomics
            45 w. Gude Dr., Rockville, MD 20850, USA
            Tel: 2404533151
            Fax: 2404534580
            Email: HoltRA@celera.com
            Plate: NU01004ABW  row: G  column: 18
            Seq primer: M13 Reverse.
FEATURES             Location/Qualifiers
     source          1. .684
                     /organism="Anopheles gambiae"
                     /strain="RSP-ST (Reduced susc. to Permethrin - std.
                     chromosome)"
                     /db_xref="taxon:7165"
                     /clone="19600449628524"
                     /clone_lib="A.Gam.ad.cDNA1"
                     /dev_stage="Adult"
                     /lab_host="DH10b"
                     /note="Vector: pSport1; Site_1: Sal1; Site_2: Not1; Whole
                     adult mosquitoes (mixed sex) frozen on liquid nitrogen.
                     cDNA inserts >500 bp cloned directionally into pSport 1.
                     Not 1 site is 3'. Clones available through the Malaria
                     Research and Reference Reagent Resource Center
                     (www.malaria.mr4.org)."
BASE COUNT      122 a    273 c    153 g    136 t
ORIGIN

  Query Match             17.6%;  Score 43.4;  DB 13;  Length 684;
  Best Local Similarity   49.3%;  Pred. No. 1.4;
  Matches  113;  Conservative    0;  Mismatches  116;  Indels    0;  Gaps    0;

Qy          1 CTGGCCCATCAGACGTCGGCCGACCAACGGGCCGCATTGCTCGAGCTCGTACGAGACCAT 60
              | || |  | ||||    |||| |||   |||| |  |  |||   ||   |  | ||
Db        309 CCGGTCGTTAAGACTGTTGCCGTCCACTCGGCCCCGGTCGTCGCCGTCCACCACGCCCCG 368

Qy         61 GTGGCGGCAGTGCTCCGGCACGCGGACCCGAAAGCCATCGCGCCCGACCAGTCGTTCCGT 120
              || |  |   |||| |  || | |||       | |||  | | ||    || |||
Db        369 GTCGTGAAGGCCCTCCCGGTCGTGCACCACGCTCCGGTCGTCCACCACGCCCCGGTCCTG 428

Qy        121 GCACTCGGCTTCGATTCACTCACGGCCGTCGAGTTCCGAAACCTGCTGATCAAGGCAACA 180
              || | || | ||   | ||| |||| | | |   ||| | |      || |
Db        429 AAGACCGTCGTCCACTCTGCCCCGGTCGTCCACTCCGTGCACCCGGTCCCGGTGGTCAAG 488

Qy        181 GGACTCCGCCTTCCTGTCTCGCTGGTCTTCGACCACCCGACCCCTGCCA 229
                |  | |   || |   | || |  || || |||||   | || | |||
Db        489 AGCGTGCATGTTGCCAGCCCGGTCCTCGTCCACCACTAAAACCGTCCCA 537


Search completed: June 18, 2003, 01:00:20
Job time : 318.438 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 17, 2003, 11:26:44 ; Search time 2502.34 Seconds
(without alignments)
15003.005 Million cell updates/sec

Title: US-09-914-286-1_COPY_6256_7545
Perfect score: 1290
Sequence: 1 gagccgatcgccatcgttgg.....tcctcgaagaagcacccgcc 1290

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1290	100.0	11096	1	AF275943	AF275943 Streptomy
2	1290	100.0	12381	6	AX006889	AX006889 Sequence
3	1290	100.0	30690	6	E38020	E38020 Avermectin
4	1290	100.0	64957	1	AB032367	AB032367 Streptomy
5	1032.2	80.0	31422	6	E38021	E38021 Avermectin
c 6	1032.2	80.0	64957	1	AB032367	AB032367 Streptomy
7	784.4	60.8	78210	1	AB070949	AB070949 Streptomy
c 8	716.6	55.6	84985	1	SNA278573	AJ278573 Streptomy
9	716	55.5	65140	6	AX211705	AX211705 Sequence
10	716	55.5	123580	1	AF263912	AF263912 Streptomy
11	716	55.5	125401	6	AX211739	AX211739 Sequence
12	715.4	55.5	37948	1	AF079138	AF079138 Streptomy
13	715.2	55.4	9513	6	AX089460	AX089460 Sequence
14	715.2	55.4	50000	6	AX089417	AX089417 Sequence
15	715.2	55.4	50000	6	AX089420	AX089420 Sequence
16	715.2	55.4	80161	1	AY007564	AY007564 Saccharop
17	715.2	55.4	80161	6	AR165018	AR165018 Sequence
c 18	714.6	55.4	26195	1	SC1G7	AL591083 Streptomy
19	713.6	55.3	104326	1	AB070940	AB070940 Streptomy
20	713.4	55.3	30000	6	AX250261	AX250261 Sequence
21	708.2	54.9	41097	1	AF016585	AF016585 Streptomy
22	708	54.9	113193	1	AF357202	AF357202 Streptomy
c 23	706	54.7	104326	1	AB070940	AB070940 Streptomy
c 24	703.8	54.6	24225	1	SC2C4	AL512902 Streptomy
c 25	696.4	54.0	30000	6	AX250263	AX250263 Sequence
26	696	54.0	20394	1	SNA132222	AJ132222 Streptomy
27	696	54.0	20394	6	AX067996	AX067996 Sequence
28	688.4	53.4	43280	1	SFU78289	U78289 Streptomyce
29	686.2	53.2	30000	6	AX250262	AX250262 Sequence
30	684.6	53.1	20235	1	SERERYAB	M63677 S.erythraea
31	684.6	53.1	20235	6	AR049368	AR049368 Sequence
32	684.6	53.1	20235	6	AR095529	AR095529 Sequence
33	684.6	53.1	20444	1	SEERYABS	X62569 S.erythraea
34	683.6	53.0	14775	6	AX089462	AX089462 Sequence
35	683	52.9	3900	1	AF262754	AF262754 Amycolato
36	678	52.6	27541	6	AX211706	AX211706 Sequence
37	677.2	52.5	30000	6	AX250263	AX250263 Sequence
38	676	52.4	32870	1	AF007101	AF007101 Streptomy
39	675.2	52.3	5544	1	SNA132221	AJ132221 Streptomy
40	675.2	52.3	5544	6	AX067994	AX067994 Sequence
41	675.2	52.3	84985	1	SNA278573	AJ278573 Streptomy
42	674.2	52.3	53784	1	AMM223012	AJ223012 Amycolato
43	674.2	52.3	53789	6	A69720	A69720 Sequence 3
44	674.2	52.3	90445	1	AF040570	AF040570 Amycolato
45	673	52.2	24568	1	AF220951	AF220951 Streptomy

ALIGNMENTS

RESULT 1
AF275943
LOCUS AF275943 11096 bp DNA linear BCT 02-SEP-2000
DEFINITION Streptomyces avermitilis avermectin polyketide synthase gene, partial cds.
ACCESSION AF275943
VERSION AF275943.1 GI:9964075
KEYWORDS .
SOURCE Streptomyces avermitilis.
ORGANISM Streptomyces avermitilis
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae; Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE 1 (bases 1 to 11096)
AUTHORS Hong,Y.-S. and Lee,J.J.
TITLE Targeted Gene Disruption of the avermectin O-methyltransferase gene

```
          and polyketide synthase gene from Streptomyces avermitilis
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 11096)
  AUTHORS   Hong,Y.-S. and Lee,J.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-JUN-2000) Anticancer Agent Research Laboratory, Korea
            Research Institute of Bioscience and Biotechnology, P.O. Box 116,
            YuSong-Gu, Taejon 305-600, South Korea
FEATURES             Location/Qualifiers
     source          1. .11096
                     /organism="Streptomyces avermitilis"
                     /strain="ATCC31271"
                     /db_xref="ATCC:31271"
                     /db_xref="taxon:33903"
     CDS             220. .>11096
                     /note="Avr1"
                     /codon_start=1
                     /transl_table=11
                     /product="avermectin polyketide synthase"
                     /protein_id="AAG09812.1"
                     /db_xref="GI:9964076"
                     /translation="MADEADGGVVFVFPGQGPQWPGMGRELLDASDVFRESVRACEAE
                     FAPYVDWSVEQVLRDSPDAPGLDRVDVVQRTLFAVMISLAALWRSQGVEPCAVLGHSL
                     GEIAAAHVSGGLSLADAARVGDAWSQAQTTLAGTGALVSVAATPDELLPRIAPWTEDN
                     PARLAVAAVNGPRSTVVSGAREAVADLVADLTAAQVRTRMIPVDVPAHSPLMYAIEER
                     VVSGLLPITPRPSRIPFHSSVTGGRLDTRELDAAYWYRNMSSTVRFEPAARLLLQQGP
                     KTFVEMSPHPVLTMGLQELAADLGDTTGTADTVIMGTLRRGQGTLDHFLTSLAQLRGH
                     GETSATTVLSARLTALSPTQQQSLLLDLVRAHTMAVLNDDGNERTASDAGPSASFAHL
                     GFDSVMGVELRNRLSKATGLRLPVTLILDHTTPAAVAARLRTAALGHLDEDTAPVPDS
                     PSGHGGTAAADDPIAIIGMACRFPGGVRSPKDLWELAASGGDAIGPFPTDRGWPTEQR
                     HAQDPTQPGTFYPQGGGFLHDAAHFDAGFFGISPREALAMDPQQRLLLETSWEAFERA
                     GIDPLSVRGSRTGVFAGALSFDYGPRMDTASSEGAADVEGHILTGTTGSVLSGRSAYS
                     FGLEGPAITVDTGCSASLVTLHLACQSLRSGECTFALAGGVSDVHPGMFIEFSRQCGL
                     SVDGRCKAVSAAADGTGWGEGIGLLAVVRGSAVNQDGASNGLTAPNGPAQERVIRQAL
                     ANAGLSVADVDVVEGHGTGTTLGDPIEAQALLATYGQRAGDRPLWLGSLKSNIGHTMA
                     AAGVGGVIKMVMALREGVLPRTLHVDEPSPQGLAAGAVRLLTEAVPWPGDAAGRLRRA
                     GVSSFGIGGTNAHVILEEAPAAGGCVAGGRVLEGAPGLAISVAESVAAPVAVSAPVAE
                     SVPVPVPVPVPVPVSARSEAGLRAQAEALRQYEAVQPDVSLADVGAGLACRQAVLEHH
                     VVILAACTSSSRSAARTTARSSSTARPQARPAPTSARRIGALAAGSGSAALTTGHAPG
                     GDRGGVVFVFPWQGGQWAGMGVRLLCLLRVFARRMQACEEALAPWVDWSVVDILRRDA
                     GDAVWEQADVVQPVLFSVMVSLAALWRSYGIEPNEVLGHSKDEIAAAHIYGALSLKDA
                     AKTVALPPQEVEQLIGERGGRLWVAAVNGPRSTAVSGDAEAVVEVLAYCAGTGVRPRI
                     PVDYASHCPHVQPLREELLELLGDISPQPYGVPFFSTVEGTWLDTTTLDAAYWYRNLH
                     QPVRFSHDVQALADQGHRVLLEVSPHPTLVPAIEDTTEDTARRRHCDRQPPPRRERHP
                     LLPQRLRLDPYYRHRQTHHVAPLLDPPRHLPPPLDAPRPAHLPFQHQHYWLESSQPGA
                     GSGSGAGAGSGAGSGRAGTAGGTAEVESRFWDAVARQDLETVATTLAVPPSAGLDTVV
                     PALSAWHRHQHDQARINTWTYQETWKPLTLPTTHQPHQTWLIAIPETQTHHPHITNIL
                     TNLHHHGITPIPLPLTTHHTNPQHLHHTLHHTRQQAQNHTTGAITGLLSLLALDETPH
                     PHHPHTPTGTLLNLTLTQTHTQTHPPTPLWYATTNATTTHPNDPLTHPTQAQTWGLAR
                     TTLLEHPTHTAGIIDLPTTPTPHTLHHLTQTLTQPHHQTQLAIRTTGTHTRRLTPTTL
                     TPTHQPPTPTPHGTTLITGGTGALATHLTHHLTTHQPTQHLLLTSRTGPHTPHAQHLT
                     TQLQQKGIHLTITTCDTSTPRPTHNNSLNTIPPQHPVTTVIHTGGILDDATLTNLTPT
                     QLNNVLRAKAHSAHLLHQLTQHTPLTAFVLYSSAAATFGAPGQANYAAANAYLDALAH
                     HRHTHHLPATSIAWGTWQGNGLADSDKARAYLDRRGFRPMSPELATAAVTQAIADTER
                     PYVVIADIDWSKIEHTSQTSDLVSAARAREPAVQRPTPPAELHKTLAHQTSADQRAAL
                     LELVRDHVAAVLRHADPKAIAPDQSFRALGFDSLTAVEFRNLLIKATGLRLPVSLVFD
                     HPTPAKLAVHLQNQLRGTAAESAPSAAAVTAEASVTEPIAIVGMACRFPGGVTSADDF
                     WDLISSEQDAIGGFPTDRGWDLDTLYDPDPDHPGTCYTRNGGFLYDAGHFDAEFFGIS
                     PREALAMDPQQRLLLETAWETIEHAGINPHTLHGTPTGVFTGTNGQDYALRVHNAGQS
                     TDGFALTGTAGSVISGRISYTFGFEGPAVSVDTACSSSLVALHLACQALRAGECSMAL
                     AGGVTVMSSPGAFVEFSRQRGLAADGHCKAFSAAADGTGWGEGVGMLLVERLSDAHRN
                     GHRVLAVVRGSAVNQDGASNGLTAPNGPSQQRVIRQALANAGLSAGDVDAVEAHGTGT
                     TLGDPIEAQALLATYGQDRAGEGPLWLGSVKSNVGHTQAAAGVAGVIKMVMALRHGLL
                     PRTLHVDEPSPHVDWSAGAVQLLTETVPWPGGEGRLRRAGVSSFGVSGTNAHVILEEA
                     PADDVPGGPPAGEGDAGSDDEAAAGSPGVWPWLVSAKSQPALRAQAQALHAHLTDHPG
                     LDLADVGYTLAHARAVFDHRATLIAADRDTFLQALQALAAGEPHPAVIHSSAPGGTGT
                     GEAAGKTAFICSGQGTQRPGMAHGLYHTHPVFAAALNDICTHLDPHLDHPLLPLLTQN
                     DNDNEDAAALLQQTRYAQPALFAFQVALHRLLTDGYHITPHYYAGHSLGEITGAHLAG
                     ILTLTDATTLITQRATLMQTMPPGTMTTLHTTPHHITHHLTAHENDLAIAAINTPTSL
                     VISGTPHTVQHITTLCQQQGIKTKTLPTNHAFHSPHTNPILNQLHQHTQTLTYHPPHT
                     PLITANTPPDQLLTPHYWTQQARNTVDYATTTQTLHQHGVTTYIELGPDNTLTTLTHH
                     NLPNTPTTTLTLTHPHHHPQTHLLTNLAKTTTTWHPHHYTHHAQPTPHPHPLDLPTYP
                     FQHHHYWLESTQPGAGNVSAAGLDPTEHPLLGATLELATDGGALLAGRLSLRSHPWLA
                     DHAVGGTVLLSGATFLELALHAGTYVGCDRVDELTLHAPLVVPVDGGVSVQVGVAAAD
                     GKGRRLVSVYARGGSACGGGASGGVWTCHASGVLVEAAAGGVVVDGLAGVWPPRGAV
                     AVDVDGVRDRLAGAGCVLGPVFSGLRAVWRDGGDLLAEVCLPEEAWGDAAGFGLHPAL
                     LNGVVQPLSVLLPGGTGFGKGAGFGKGVRVPAVWGGVSLHRAGVTGVRVRVSAVGRGG
                     GREAVSVVVGDEAGVPVASVDRLELRPVDMGQLRAVSVSAGRRGSLYAVQWAEVGPVP
                     VCGQAWAWHEDVGESGGGPVPGVVVLRCPDAGAGGGGGGGGGGVGEVVGGVLGVVQG
                     WLGLERFAGSRLVVVTRGAVVAGPEDGPVDVVGASVWGLVRSAQAEHPDRFVLLDLDT
                     DTGTDLDTGAGAGWGVDGGRVAAVVACGEPQLAVRGERLLAARLTRLESSGDVPAQRS
                     GDTRARRSDVPAQRSGGVPARRSVDVSGREVLPWLSGGSVLVTGGTGVLGAAVARHLA
                     GVCGVRDLLLVSRRGPDAPGAEGLRRSWPRGAEVRIVACDVGERREVVRLAGGCSCRV
                     SVFVDFSAVASVTVRLPVASDVRKEAAMAYATVEEFTDYLDPDP"
BASE COUNT     1756 a   4059 c   3528 g   1753 t
ORIGIN

  Query Match           100.0%;  Score 1290;  DB 1;  Length 11096;
  Best Local Similarity  100.0%;  Pred. No. 1e-150;
  Matches 1290;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GAGCCGATCGCCATCGTTGGCATGGCCTGTCGTTTCCCCGGCGGAGTGACCTCGGCGGAC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6376 GAGCCGATCGCCATCGTTGGCATGGCCTGTCGTTTCCCCGGCGGAGTGACCTCGGCGGAC 6435

Qy         61 GACTTCTGGGATCTGATCTCCTCCGAGCAGGACGCGATCGGCGGATTCCCCACCGACCGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6436 GACTTCTGGGATCTGATCTCCTCCGAGCAGGACGCGATCGGCGGATTCCCCACCGACCGC 6495

Qy        121 GGCTGGGACCTGGACACGCTCTACGACCCCGACCCCGACCACCCCGGCACCTGCTACACC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6496 GGCTGGGACCTGGACACGCTCTACGACCCCGACCCCGACCACCCCGGCACCTGCTACACC 6555

Qy        181 CGAAACGGCGGATTCCTCTACGACGCAGGCCACTTCGACGCCGAATTCTTCGGCATCAGC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6556 CGAAACGGCGGATTCCTCTACGACGCAGGCCACTTCGACGCCGAATTCTTCGGCATCAGC 6615

Qy        241 CCCCGCGAAGCCCTCGCCATGGACCCCCAGCAACGACTCCTCCTCGAAACCGCCTGGGAA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6616 CCCCGCGAAGCCCTCGCCATGGACCCCCAGCAACGACTCCTCCTCGAAACCGCCTGGGAA 6675

Qy        301 ACCATCGAACACGCCGGCATCAACCCCCACACCCTCCACGGCACCCCCACCGGAGTCTTC 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6676 ACCATCGAACACGCCGGCATCAACCCCCACACCCTCCACGGCACCCCCACCGGAGTCTTC 6735

Qy        361 ACCGGCACCAACGGACAGGACTACGCACTTCGCGTGCACAACGCGGGCCAGTCAACCGAT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6736 ACCGGCACCAACGGACAGGACTACGCACTTCGCGTGCACAACGCGGGCCAGTCAACCGAT 6795

Qy        421 GGTTTCGCACTGACCGGAACCGCCGGCAGCGTCATCTCCGGTCGTATCTCGTACACGTTT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6796 GGTTTCGCACTGACCGGAACCGCCGGCAGCGTCATCTCCGGTCGTATCTCGTACACGTTT 6855

Qy        481 GGTTTTGAGGGTCCTGCGGTGTCGGTGGACACGGCTTGTTCCTCGTCGTTGGTGGCTTTG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6856 GGTTTTGAGGGTCCTGCGGTGTCGGTGGACACGGCTTGTTCCTCGTCGTTGGTGGCTTTG 6915

Qy        541 CATCTGGCCTGTCAGGCGTTGCGTGCGGGTGAGTGCTCGATGGCGCTTGCCGGGGGTGTG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6916 CATCTGGCCTGTCAGGCGTTGCGTGCGGGTGAGTGCTCGATGGCGCTTGCCGGGGGTGTG 6975

Qy        601 ACGGTGATGTCGTCTCCGGGTGCCTTCGTGGAGTTTTCGCGGCAGCGGGGTCTGGCCGCG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6976 ACGGTGATGTCGTCTCCGGGTGCCTTCGTGGAGTTTTCGCGGCAGCGGGGTCTGGCCGCG 7035

Qy        661 GACGGGCATTGCAAGGCGTTCTCGGCGGCGGCGGACGGGACCGGCTGGGGTGAGGGTGTG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7036 GACGGGCATTGCAAGGCGTTCTCGGCGGCGGCGGACGGGACCGGCTGGGGTGAGGGTGTG 7095

Qy        721 GGGATGCTGCTGGTGGAGCGGCTCTCCGACGCCCATCGCAACGGTCACCGTGTCCTGGCC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7096 GGGATGCTGCTGGTGGAGCGGCTCTCCGACGCCCATCGCAACGGTCACCGTGTCCTGGCC 7155

Qy        781 GTGGTGCGTGGCAGTGCGGTCAACCAGGACGGTGCGAGCAACGGTCTGACCGCGCCCAAC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7156 GTGGTGCGTGGCAGTGCGGTCAACCAGGACGGTGCGAGCAACGGTCTGACCGCGCCCAAC 7215
```

```
Qy       841 GGGCCGTCCCAGCAGCGTGTCATCCGCCAGGCCCTCGCCAACGCCGGCTTGTCGGCCGGT 900
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      7216 GGGCCGTCCCAGCAGCGTGTCATCCGCCAGGCCCTCGCCAACGCCGGCTTGTCGGCCGGT 7275

Qy       901 GATGTCGACGCGGTGGAGGCCCACGGCACCGGCACCACTTTGGGCGACCCGATCGAGGCC 960
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      7276 GATGTCGACGCGGTGGAGGCCCACGGCACCGGCACCACTTTGGGCGACCCGATCGAGGCC 7335

Qy       961 CAGGCCCTCCTCGCGACCTACGGACAGGACCGTGCCGGCGAGGGGCCGCTGTGGCTGGGC 1020
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      7336 CAGGCCCTCCTCGCGACCTACGGACAGGACCGTGCCGGCGAGGGGCCGCTGTGGCTGGGC 7395

Qy      1021 TCGGTCAAGTCCAATGTCGGTCACACACAGGCTGCCGCGGGCGTCGCCGGGGTGATCAAG 1080
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      7396 TCGGTCAAGTCCAATGTCGGTCACACACAGGCTGCCGCGGGCGTCGCCGGGGTGATCAAG 7455

Qy      1081 ATGGTGATGGCGCTGCGGCATGGTCTGCTGCCGCGGACGTTGCATGTGGATGAGCCGTCG 1140
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      7456 ATGGTGATGGCGCTGCGGCATGGTCTGCTGCCGCGGACGTTGCATGTGGATGAGCCGTCG 7515

Qy      1141 CCGCATGTGGACTGGTCCGCGGGTGCGGTGCAGCTGCTGACGGAGACGGTGCCCTGGCCC 1200
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      7516 CCGCATGTGGACTGGTCCGCGGGTGCGGTGCAGCTGCTGACGGAGACGGTGCCCTGGCCC 7575

Qy      1201 GGCGGGGAGGGGCGGCTACGGCGGGCAGGAGTGTCATCATTCGGCGTCAGCGGCACCAAC 1260
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      7576 GGCGGGGAGGGGCGGCTACGGCGGGCAGGAGTGTCATCATTCGGCGTCAGCGGCACCAAC 7635

Qy      1261 GCCCACGTCATCCTCGAAGAAGCACCCGCC 1290
            |||||||||||||||||||||||||||||
Db      7636 GCCCACGTCATCCTCGAAGAAGCACCCGCC 7665
```

RESULT 2
AX006889
LOCUS AX006889 12381 bp DNA linear PAT 06-SEP-2000
DEFINITION Sequence 1 from Patent WO0001827.
ACCESSION AX006889
VERSION AX006889.1 GI:9994904
KEYWORDS .
SOURCE Streptomyces avermitilis.
ORGANISM Streptomyces avermitilis
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE 1 (bases 1 to 12381)
AUTHORS Kellenberger,J.L., Leadlay,P.F., Staunton,J., Mcarthur,H.A. and
Stutzman-Engwall,K.J.
TITLE Polyketides, their preparation, and materials for use therein
JOURNAL Patent: WO 0001827-A 1 13-JAN-2000;
KELLENBERGER JOHANNES LAURENZ (CH); LEADLAY PETER FRANCIS (GB);
STAUNTON JAMES (GB); BIOTICA TECH LTD (GB); PFIZER (US); MCARTHUR
HAMISH ALASTAIR IRVIN (US); STUTZMAN ENGWALL KIM JONELLE (US)
FEATURES Location/Qualifiers
source 1. .12381
/organism="Streptomyces avermitilis"
/db_xref="taxon:33903"
BASE COUNT 1884 a 4561 c 4005 g 1931 t
ORIGIN

Query Match 100.0%; Score 1290; DB 6; Length 12381;
Best Local Similarity 100.0%; Pred. No. 1e-150;
Matches 1290; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 GAGCCGATCGCCATCGTTGGCATGGCCTGTCGTTTCCCCGGCGGAGTGACCTCGGCGGAC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7073 GAGCCGATCGCCATCGTTGGCATGGCCTGTCGTTTCCCCGGCGGAGTGACCTCGGCGGAC 7132

Qy         61 GACTTCTGGGATCTGATCTCCTCCGAGCAGGACGCGATCGGCGGATTCCCCACCGACCGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7133 GACTTCTGGGATCTGATCTCCTCCGAGCAGGACGCGATCGGCGGATTCCCCACCGACCGC 7192

Qy        121 GGCTGGGACCTGGACACGCTCTACGACCCCGACCCCGACCACCCCGGCACCTGCTACACC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7193 GGCTGGGACCTGGACACGCTCTACGACCCCGACCCCGACCACCCCGGCACCTGCTACACC 7252

Qy        181 CGAAACGGCGGATTCCTCTACGACGCAGGCCACTTCGACGCCGAATTCTTCGGCATCAGC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7253 CGAAACGGCGGATTCCTCTACGACGCAGGCCACTTCGACGCCGAATTCTTCGGCATCAGC 7312

Qy        241 CCCCGCGAAGCCCTCGCCATGGACCCCCAGCAACGACTCCTCCTCGAAACCGCCTGGGAA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7313 CCCCGCGAAGCCCTCGCCATGGACCCCCAGCAACGACTCCTCCTCGAAACCGCCTGGGAA 7372

Qy        301 ACCATCGAACACGCCGGCATCAACCCCCACACCCTCCACGGCACCCCCACCGGAGTCTTC 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7373 ACCATCGAACACGCCGGCATCAACCCCCACACCCTCCACGGCACCCCCACCGGAGTCTTC 7432

Qy        361 ACCGGCACCAACGGACAGGACTACGCACTTCGCGTGCACAACGCGGGCCAGTCAACCGAT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7433 ACCGGCACCAACGGACAGGACTACGCACTTCGCGTGCACAACGCGGGCCAGTCAACCGAT 7492

Qy        421 GGTTTCGCACTGACCGGAACCGCCGGCAGCGTCATCTCCGGTCGTATCTCGTACACGTTT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7493 GGTTTCGCACTGACCGGAACCGCCGGCAGCGTCATCTCCGGTCGTATCTCGTACACGTTT 7552

Qy        481 GGTTTTGAGGGTCCTGCGGTGTCGGTGGACACGGCTTGTTCCTCGTCGTTGGTGGCTTTG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7553 GGTTTTGAGGGTCCTGCGGTGTCGGTGGACACGGCTTGTTCCTCGTCGTTGGTGGCTTTG 7612

Qy        541 CATCTGGCCTGTCAGGCGTTGCGTGCGGGTGAGTGCTCGATGGCGCTTGCCGGGGGTGTG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7613 CATCTGGCCTGTCAGGCGTTGCGTGCGGGTGAGTGCTCGATGGCGCTTGCCGGGGGTGTG 7672

Qy        601 ACGGTGATGTCGTCTCCGGGTGCCTTCGTGGAGTTTTCGCGGCAGCGGGGTCTGGCCGCG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7673 ACGGTGATGTCGTCTCCGGGTGCCTTCGTGGAGTTTTCGCGGCAGCGGGGTCTGGCCGCG 7732

Qy        661 GACGGGCATTGCAAGGCGTTCTCGGCGGCGGCGGACGGGACCGGCTGGGGTGAGGGTGTG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7733 GACGGGCATTGCAAGGCGTTCTCGGCGGCGGCGGACGGGACCGGCTGGGGTGAGGGTGTG 7792

Qy        721 GGGATGCTGCTGGTGGAGCGGCTCTCCGACGCCCATCGCAACGGTCACCGTGTCCTGGCC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7793 GGGATGCTGCTGGTGGAGCGGCTCTCCGACGCCCATCGCAACGGTCACCGTGTCCTGGCC 7852

Qy        781 GTGGTGCGTGGCAGTGCGGTCAACCAGGACGGTGCGAGCAACGGTCTGACCGCGCCCAAC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7853 GTGGTGCGTGGCAGTGCGGTCAACCAGGACGGTGCGAGCAACGGTCTGACCGCGCCCAAC 7912

Qy        841 GGGCCGTCCCAGCAGCGTGTCATCCGCCAGGCCCTCGCCAACGCCGGCTTGTCGGCCGGT 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7913 GGGCCGTCCCAGCAGCGTGTCATCCGCCAGGCCCTCGCCAACGCCGGCTTGTCGGCCGGT 7972

Qy        901 GATGTCGACGCGGTGGAGGCCCACGGCACCGGCACCACTTTGGGCGACCCGATCGAGGCC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7973 GATGTCGACGCGGTGGAGGCCCACGGCACCGGCACCACTTTGGGCGACCCGATCGAGGCC 8032

Qy        961 CAGGCCCTCCTCGCGACCTACGGACAGGACCGTGCCGGCGAGGGGCCGCTGTGGCTGGGC 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8033 CAGGCCCTCCTCGCGACCTACGGACAGGACCGTGCCGGCGAGGGGCCGCTGTGGCTGGGC 8092

Qy       1021 TCGGTCAAGTCCAATGTCGGTCACACACAGGCTGCCGCGGGCGTCGCCGGGGTGATCAAG 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8093 TCGGTCAAGTCCAATGTCGGTCACACACAGGCTGCCGCGGGCGTCGCCGGGGTGATCAAG 8152

Qy       1081 ATGGTGATGGCGCTGCGGCATGGTCTGCTGCCGCGGACGTTGCATGTGGATGAGCCGTCG 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8153 ATGGTGATGGCGCTGCGGCATGGTCTGCTGCCGCGGACGTTGCATGTGGATGAGCCGTCG 8212

Qy       1141 CCGCATGTGGACTGGTCCGCGGGTGCGGTGCAGCTGCTGACGGAGACGGTGCCCTGGCCC 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8213 CCGCATGTGGACTGGTCCGCGGGTGCGGTGCAGCTGCTGACGGAGACGGTGCCCTGGCCC 8272

Qy       1201 GGCGGGGAGGGGCGGCTACGGCGGGCAGGAGTGTCATCATTCGGCGTCAGCGGCACCAAC 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db       8273 GGCGGGGAGGGGCGGCTACGGCGGGCAGGAGTGTCATCATTCGGCGTCAGCGGCACCAAC 8332

Qy       1261 GCCCACGTCATCCTCGAAGAAGCACCCGCC 1290
              |||||||||||||||||||||||||||||
Db       8333 GCCCACGTCATCCTCGAAGAAGCACCCGCC 8362


RESULT 3
E38020
LOCUS       E38020                30690 bp    DNA     linear   PAT 31-JAN-2002
DEFINITION  Avermectin aglycon synthase gene.
ACCESSION   E38020
VERSION     E38020.1  GI:18626909
KEYWORDS    JP 2000245457-A/1.
SOURCE      Streptomyces avermitilis.
  ORGANISM  Streptomyces avermitilis
            Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
            Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1  (bases 1 to 30690)
  AUTHORS   Omura,S. and Ikeda,H.
  TITLE     Avermectin aglycon synthase gene
  JOURNAL   Patent: JP 2000245457-A 1 12-SEP-2000;
            THE KITASATO INSTITUTE
COMMENT     OS   Streptomyces avermitilis
            PN   JP 2000245457-A/1
            PD   12-SEP-2000
            PF   24-FEB-1999 JP 1999046961
            PR
            PI   SATOSHI OMURA,HARUO IKEDA
            PC   C12N15/00,A61K31/70,C12N1/15,C12N9/88,C12P19/62, PC
            C12Q1/68//C07D493/22,
            PC   (C12N1/15,C12R1:465),(C12N9/88,C12R1:465),C12N15/00 CC
            FH   Key             Location/Qualifiers
            FT   CDS             (1). .(11916)
            FT   CDS             (11971). .(30687).
FEATURES             Location/Qualifiers
     source          1. .30690
                     /organism="Streptomyces avermitilis"
                     /db_xref="taxon:33903"
BASE COUNT     5355 a  12455 c   8617 g   4263 t
ORIGIN

  Query Match           100.0%;  Score 1290;  DB 6;  Length 30690;
  Best Local Similarity  100.0%;  Pred. No. 8e-151;
  Matches 1290;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GAGCCGATCGCCATCGTTGGCATGGCCTGTCGTTTCCCCGGCGGAGTGACCTCGGCGGAC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6256 GAGCCGATCGCCATCGTTGGCATGGCCTGTCGTTTCCCCGGCGGAGTGACCTCGGCGGAC 6315

Qy         61 GACTTCTGGGATCTGATCTCCTCCGAGCAGGACGCGATCGGCGGATTCCCCACCGACCGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6316 GACTTCTGGGATCTGATCTCCTCCGAGCAGGACGCGATCGGCGGATTCCCCACCGACCGC 6375

Qy        121 GGCTGGGACCTGGACACGCTCTACGACCCCGACCCCGACCACCCCGGCACCTGCTACACC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6376 GGCTGGGACCTGGACACGCTCTACGACCCCGACCCCGACCACCCCGGCACCTGCTACACC 6435

Qy        181 CGAAACGGCGGATTCCTCTACGACGCAGGCCACTTCGACGCCGAATTCTTCGGCATCAGC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6436 CGAAACGGCGGATTCCTCTACGACGCAGGCCACTTCGACGCCGAATTCTTCGGCATCAGC 6495

Qy        241 CCCCGCGAAGCCCTCGCCATGGACCCCCAGCAACGACTCCTCCTCGAAACCGCCTGGGAA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6496 CCCCGCGAAGCCCTCGCCATGGACCCCCAGCAACGACTCCTCCTCGAAACCGCCTGGGAA 6555

Qy        301 ACCATCGAACACGCCGGCATCAACCCCCACACCCTCCACGGCACCCCCACCGGAGTCTTC 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6556 ACCATCGAACACGCCGGCATCAACCCCCACACCCTCCACGGCACCCCCACCGGAGTCTTC 6615

Qy        361 ACCGGCACCAACGGACAGGACTACGCACTTCGCGTGCACAACGCGGGCCAGTCAACCGAT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6616 ACCGGCACCAACGGACAGGACTACGCACTTCGCGTGCACAACGCGGGCCAGTCAACCGAT 6675

Qy        421 GGTTTCGCACTGACCGGAACCGCCGGCAGCGTCATCTCCGGTCGTATCTCGTACACGTTT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6676 GGTTTCGCACTGACCGGAACCGCCGGCAGCGTCATCTCCGGTCGTATCTCGTACACGTTT 6735

Qy        481 GGTTTTGAGGGTCCTGCGGTGTCGGTGGACACGGCTTGTTCCTCGTCGTTGGTGGCTTTG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6736 GGTTTTGAGGGTCCTGCGGTGTCGGTGGACACGGCTTGTTCCTCGTCGTTGGTGGCTTTG 6795

Qy        541 CATCTGGCCTGTCAGGCGTTGCGTGCGGGTGAGTGCTCGATGGCGCTTGCCGGGGGTGTG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6796 CATCTGGCCTGTCAGGCGTTGCGTGCGGGTGAGTGCTCGATGGCGCTTGCCGGGGGTGTG 6855

Qy        601 ACGGTGATGTCGTCTCCGGGTGCCTTCGTGGAGTTTTCGCGGCAGCGGGGTCTGGCCGCG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6856 ACGGTGATGTCGTCTCCGGGTGCCTTCGTGGAGTTTTCGCGGCAGCGGGGTCTGGCCGCG 6915

Qy        661 GACGGGCATTGCAAGGCGTTCTCGGCGGCGGCGGACGGGACCGGCTGGGGTGAGGGTGTG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6916 GACGGGCATTGCAAGGCGTTCTCGGCGGCGGCGGACGGGACCGGCTGGGGTGAGGGTGTG 6975

Qy        721 GGGATGCTGCTGGTGGAGCGGCTCTCCGACGCCCATCGCAACGGTCACCGTGTCCTGGCC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6976 GGGATGCTGCTGGTGGAGCGGCTCTCCGACGCCCATCGCAACGGTCACCGTGTCCTGGCC 7035

Qy        781 GTGGTGCGTGGCAGTGCGGTCAACCAGGACGGTGCGAGCAACGGTCTGACCGCGCCCCAAC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7036 GTGGTGCGTGGCAGTGCGGTCAACCAGGACGGTGCGAGCAACGGTCTGACCGCGCCCAAC 7095

Qy        841 GGGCCGTCCCAGCAGCGTGTCATCCGCCAGGCCCTCGCCAACGCCGGCTTGTCGGCCGGT 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7096 GGGCCGTCCCAGCAGCGTGTCATCCGCCAGGCCCTCGCCAACGCCGGCTTGTCGGCCGGT 7155

Qy        901 GATGTCGACGCGGTGGAGGCCCACGGCACCGGCACCACTTTGGGCGACCCGATCGAGGCC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7156 GATGTCGACGCGGTGGAGGCCCACGGCACCGGCACCACTTTGGGCGACCCGATCGAGGCC 7215

Qy        961 CAGGCCCTCCTCGCGACCTACGGACAGGACCGTGCCGGCGAGGGGCCGCTGTGGCTGGGC 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7216 CAGGCCCTCCTCGCGACCTACGGACAGGACCGTGCCGGCGAGGGGCCGCTGTGGCTGGGC 7275

Qy       1021 TCGGTCAAGTCCAATGTCGGTCACACACAGGCTGCCGCGGGCGTCGCCGGGGTGATCAAG 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7276 TCGGTCAAGTCCAATGTCGGTCACACACAGGCTGCCGCGGGCGTCGCCGGGGTGATCAAG 7335

Qy       1081 ATGGTGATGGCGCTGCGGCATGGTCTGCTGCCGCGGACGTTGCATGTGGATGAGCCGTCG 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7336 ATGGTGATGGCGCTGCGGCATGGTCTGCTGCCGCGGACGTTGCATGTGGATGAGCCGTCG 7395

Qy       1141 CCGCATGTGGACTGGTCCGCGGGTGCGGTGCAGCTGCTGACGGAGACGGTGCCCTGGCCC 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7396 CCGCATGTGGACTGGTCCGCGGGTGCGGTGCAGCTGCTGACGGAGACGGTGCCCTGGCCC 7455

Qy       1201 GGCGGGGAGGGGCGGCTACGGCGGGCAGGAGTGTCATCATTCGGCGTCAGCGGCACCAAC 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7456 GGCGGGGAGGGGCGGCTACGGCGGGCAGGAGTGTCATCATTCGGCGTCAGCGGCACCAAC 7515

Qy       1261 GCCCACGTCATCCTCGAAGAAGCACCCGCC 1290
              |||||||||||||||||||||||||||||
Db       7516 GCCCACGTCATCCTCGAAGAAGCACCCGCC 7545


RESULT 4
AB032367
LOCUS       AB032367              64957 bp    DNA     linear   BCT 18-SEP-1999
DEFINITION  Streptomyces avermitilis polyketide synthase gene cluster (aveA1,
            aveA2, aveA3, aveA4) and aveC, aveE genes, complete cds.
ACCESSION   AB032367
VERSION     AB032367.1  GI:5902890
KEYWORDS    AveC; cytochrome P450 hydroxylase; type I polyketide synthase AVES
            4; type I polyketide synthase AVES 3; type I polyketide synthase
            AVES 2; type I polyketide synthase AVES 1.
SOURCE      Streptomyces avermitilis DNA.
```

```
  ORGANISM  Streptomyces avermitilis
            Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
            Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1  (bases 1 to 64957)
  AUTHORS   Ikeda,H., Nonomiya,T., Usami,M., Ohta,T. and Omura,S.
  TITLE     Organization of the biosynthetic gene cluster for the polyketide
            anthelmintic macrolide avermectin in Streptomyces avermitilis
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9509-9514 (1999)
  MEDLINE   99380548
REFERENCE   2  (bases 1 to 64957)
  AUTHORS   Ikeda,H., Nonomiya,T., Usami,M., Ohta,T. and Omura,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-SEP-1999) Haruo Ikeda, School of Pharmaceutical
            Sciences, Kitasato University, Microbial chemistry; 5-9-1
            Shirokane, Minato-ku, Tokyo 108-8641, Japan
            (E-mail:ikeda@mc.pharm.kitasato-u.ac.jp, Tel:+81-3-5791-6242,
            Fax:+81-3-3444-6197)
FEATURES            Location/Qualifiers
     source         1. .64957
                    /organism="Streptomyces avermitilis"
                    /db_xref="taxon:33903"
     gene           101. .12019
                    /gene="aveA1"
     CDS            101. .12019
                    /gene="aveA1"
                    /note="contains a loading module, Ave module 1 and module
                    2
                    multifunctional polyketide synthase"
                    /codon_start=1
                    /transl_table=11
                    /product="type I polyketide synthase AVES 1"
                    /protein_id="BAA84474.1"
                    /db_xref="GI:5902891"
                    /translation="MQRMDGGEEPRPAAGEVLGVADEADGGVVFVFPGQGPQWPGMGR
                    ELLDASDVFRESVRACEAAFAPYVDWSVEQVLRDSPDAPGLDRVDVVQPTLFAVMISL
                    AALWRSQGVEPCAVLGHSLGEIAAAHVSGGLSLADAARVVTLWSQAQTTLAGTGALVS
                    VAATPDELLPRIAPWTEDNPARLAVAAVNGPRSTVVSGAREAVADLVADLTAAQVRTR
                    MIPVDVPAHSPLMYAIEERVVSGLLPITPRPSRIPFHSSVTGGRLDTRELDAAYWYRN
                    MSSTVRFEPAARLLLQQGPKTFVEMSPHPVLTMGLQELAPDLGDTTGTADTVIMGTLR
                    RGQGTLDHFLTSLAQLRGHGETSATTVLSARLTALSPTQQQSLLLDLVRAHTMAVLND
                    DGNERTASDAGPSASFAHLGFDSVMGVELRNRLSKATGLRLPVTLIFDHTTPAAVAAR
                    LRTAALGHLDEDTAPVPDSPSGHGGTAAADDPIAIIGMACRFPGGVRSPKDLWELAAS
                    GGDAIGPFPTDRGWPTEQRHAQDPTQPGTFYPQGGGFLHDAAHFDAGFFGISPREALA
                    MDPQQRLLLETSWEAFERAGIDPLSVRGSRTGVFAGALSFDYGPRMDTASSEGAADVE
                    GHILTGTTGSVLSGRIAYSFGLEGPAITVDTGCSASLVTLHLACQSLRSGECTLALAG
                    GVSVMSTLGMFIEFSRQRGLSVDGRCKAYSAAADGTGWGEGVGMLLVERLSDAVRLGH
                    RVLAVVRGSAVNQDGASNGLTAPNGPAQERVIRQALANAGLSVADVDVVEGHGTGTTL
                    GDPIEAQALLATYGQRAGDRPLWLGSLKSNIGHTMAAAGVGGVIKMVMALREGVLPRT
                    LHVDKPSPQVDWSAGAVRLLTEAVPWPGDAAGRLRRAGVSSFGIGGTNAHVILEEAPA
                    AGGCVAGGGVLEGAPGLAISVAESVAAPVAVSAPVAESVPVPVPVPVPVSARSEAG
                    LRAQAEALRQYVAVRPDVSLADVGAGLACGRAVLEHRAVVLAADREELVQGLGALAAG
                    EPDRRVTTGHAPGGDRGGVVFVFPGQGGQWAGMGVRLLASSPVFARRMQACEEALAPW
                    VDWSVVDILRRDAGDAVWERADVVQPVLFSVMVSLAALWRSYGIEPDAVLGHSQGEIA
                    AAHVCGALSLKDAAKTVALRSRALAAVRGRGGMASVPLPAQEVEQLIGERWAGRLWVA
                    AVNGPRSTAVSGDAEAVDEVLAYCAGTGVRARRIPVDYASHCPHVQPLREELLELLGD
                    ISPQPSGVPFFSTVEGTWLDTTTLDAAYWYRNLHQPVRFSDAVQALADDGHRVFVEVS
                    PHPTLVPAIEDTTEDTAEDVTAIGSLRRGDNDTRRFLTALAHTHTTGIGTPTTWHHHY
                    THHHTHPHPHTHLDLPTYPFQHQHYWLESSQPGAGSGSGAGAGSGAGSGRAGTAGGTA
                    EVESRFWDAVARQDLETVATTLAVPPSAGLDTVVPALSAWHRHQHDQARINTWTYQET
                    WKPLTLPTTHQPHQTWLIAIPETQTHHPHITNILTNLHHHGITPIPLTLNHTHTNPQH
                    LHHTLHHTRQQAQNHTTGAITGLLSLLALDETPHPHHPHTPTGTLLNLTLTQTHTQTH
                    PPTPLWYATTNATTTHPNDPLTHPTQAQTWGLARTTLLEHPTHTAGIIDLPTTPTPHT
                    LQHLTQTLTQPHHQTQLAIRTTGTHTRRLTPTTLTPTHQPPTPTPHGTTLITGGTGAL
                    ATHLTHHLTTHQPTQHLLLTSRTGPHTPHAQHLTTQLQQKGIHLTITTCDTSNPDQLQ
                    QLLNTIPPQHPLTTVIHTAGILDDATLTNLTPTQLNNVLRAKAHSAHLLHQLTQHTPL
                    TAFVLYSSAAATFGAPGQANYAAANAYLDALAHHRHTHHLPATSIAWGTWQGNGLADS
                    DKARAYLDRRGFRPMSPELATAAVTQAIADTERPYVVIADIDWSKIEHTSQTSDLVSA
                    ARELEPAVQRPTPPAELHKTLAHQTSADQRAALLELVRDHVAAVLRHADPKAIAPDQS
                    FRALGFDSLTAVEFRNLLIKATGLRLPVSLVFDHPTPAKLAVHLQNQLRGTAAESAPS
                    AAAVTAEASVTEPIAIVGMACRFPGGVTSADDFWDLISSEQDAIGGFPTDRGWDLDTL
                    YDPDPDHPGTCYTRNGGFLYDAGHFDAEFFGISPREALAMDPQQRLLLETAWETIEHA
                    GINPHTLHGTPTGVFTGTNGQDYALRVHNAGQSTDGFALTGTAGSVISGRISYTFGFE
                    GPAVSVDTACSSSLVALHLACQALRAGECSMALAGGVTVMSSPGAFVEFSRQRGLAAD
                    GHCKAFSAAADGTGWGEGVGMLLVERLSDAHRNGHRVLAVVRGSAVNQDGASNGLTAP
                    NGPSQQRVIRQALANAGLSAGDVDAVEAHGTGTTLGDPIEAQALLATYGQDRAGEGPL
                    WLGSVKSNVGHTQAAAGVAGVIKMVMALRHGLLPRTLHVDEPSPHVDWSAGAVQLLTE
                    TVPWPGGEGRLRRAGVSSFGVSGTNAHVILEEAPADDVPGGPPAGEGDAGSDDEAAAG
                    SPGVWPWLVSAKSQPALRAQAQALHAHLTDHPGLDLADVGYTLAHARAVFDHRATLIA
                    ADRDTFLQALQALAAGEPHPAVIHSSAPGGTGTGEAAGKTAFICSGQGTQRPGMAHGL
                    YHTHPVFAAALNDICTHLDPHLDHPLLPLLTQNDNDNEDAAALLQQTRYAQPALFAFQ
                    VALHRLLTDGYHITPHYYAGHSLGEITAAHLAGILTLTDATTLITQRATLMQTMPPGT
                    MTTLHTTPHHITHHLTAHENDLAIAAINTPTSLVISGTPHTVQHITTLCQQQGIKTKT
                    LPTNHAFHSPHTNPILNQLHQHTQTLTYHPPHTPLITANTPPDQLLTPHYWTQQARNT
                    VDYATTTQTLHQHGVTTYIELGPDNTLTTLTHHNLPNPPTTTLTLTHPHHHPQTHLLT
                    NLAKTTTTWHPHHYTHHDNQPHTHTHLDLPTYPFQHHHYWLESTQPGAGNVSAAGLDP
                    TEHPLLGATLELATDGGALLAGRLSLRSHPWLADHAVGGTVLLSGATFLELALHAGTY
                    VGCDRVDELTLHAPLVVPVDGGVSVQVGVAAADGEGRRLVSVYARGGSACGGGGASGG
                    VWTCHASGVLVEAAAGGVVVDGLAGVWPPRGAVAVDVDGVRDRLAGAGCVLGPVFSGL
                    RAVWRDGGDLLAEVCLPEEAWGDAAGFGLHPALLDGVVQPLSVLLPGGTGFGEGAGFG
                    EGVRVPAVWGGVSLHRAGVTGVRVRVSAVGRGGGREAVSVVVGDEAGVPVASVDRLEL
                    RPVDMGQLRAVSVSAGRRGSLYAVQWAEVGPVPVCGQAWAWHEDVGESGGGPVPGVVV
                    LRCPDAGAGGGGGGGGGVGEVVGGVLGVVQGWLGLERFAGSRLVVVTRGAVVAGPE
                    DGPVDVVGASVWGLVRSAQAEHPDRFVLLDLDTDTGTDLDTGAGAGWGVDGGRVAAVV
                    ACGEPQLAVRGERLLAARLKRLESSGDVPAQRSGDTRARRSDVPAQRSGGVPARRSVD
                    VSGREVLPWLSGGSVLVTGGTGVLGAAVARHLAGVCGVRDLLLVSRRGPDAPGAEGLR
                    AELAALGAEVRIVACDVGERREVVRLLEGVPAGCPLTGVVHAAGVLDDATIASLTPER
                    LGTVFAAKVDAALLLDELTRGMELSAFVLFSSAAGILGSAGQGNYAAANAALDALAYR
                    RRAAGLPGVSLAWGLWEEASGMTGHLAGTDHRRIIRSGLHPMSTPDALALFDAALALD
                    RPVLLPADLRPAPPLPPLLQDLLPATRRRTTRTTTTGGADNGAQLHARLAGQTHEQQH
                    TTLLALVRSHIATVLGHTTPDTIPPDRAFRDLGFDSLTAVELRNRLSRTTGLRLPTTL
                    AFDHPNPTTLTHHLHTQLQPQPDNAVAPVLAELDKLESALSALDKTDSASERVTLRLK
                    SLMLRWNAPQHPTAESADDDEKFTSATEAEIFKFIDNDLGLS"
gene                12071. .30790
                    /gene="aveA2"
CDS                 12071. .30790
                    /gene="aveA2"
                    /note="contains Ave module 3, module 4, module 5 and
                    module 6
                    multifunctional polyketide synthase"
                    /codon_start=1
                    /transl_table=11
                    /product="type I polyketide synthase AVES 2"
                    /protein_id="BAA84475.1"
                    /db_xref="GI:5902892"
                    /translation="MQLANEAKLLEYLKRVTADLDRTRRRLYEVVEREQEPIAIVGMA
                    CRYPGGATSPTRLWHLVKSQTDAIGEFPTDRGWNLEQLYDPDPDRSGTSYTRSGGFLY
                    DAGDFDAAFFELSPREALAMDPQQRLLLETTWETFEQGGIDPRSMRGSRTGVFVGINP
                    EDYTTGYTHQPSNAVEGYLLTGSAASIASGRISYNFGLEGPAITIDTACSSSLVALHL
                    ACQALRSGECTMALAGGASVMATPFVFTEFSRQRGLAADGRCKAFSAAADGTGWSEGV
                    GMLLVERLSDARRNGHRVLAVVRGSAVNQDGASNGLTAPNGRSQVKVIRQALANAHLS
                    PADVDAVEAHGTGTTLGDPIEAQALVEAYGQDRPNGRPLWLGTLKSNIGHSMAAAGVG
                    GVIKMVMALRNGLLPRTLHVDEPSPHVDWSAGAVQLLTETVPWPGGEGRLRRAGVSSF
                    GVSGTNAHVILEEAPAHNIPSDTPADDVPGESAADEDAGSGDEAAAGSPGVWPWLVSA
                    KSQPALRAQAQALHAHLTDHPGLDLADVGYTLAHARAVFDHRATLIAADRDTFLQALQ
                    ALAAGEPHPAVIHSSAPGGTGTGEAAGKTAFICSGQGTQRPGMAHGLYHTHPVFAAAL
                    NDICTHLDPHLDHPLLPLLTQDPNTQDTTTLEEAAALLQQTRYAQPALFAFQVALHRL
                    LTDGYHITPHYYAGHSLGEITAAHLAGILTLTDATTLITQRATLMQTMPPGTMTTLHT
                    TPHHITHHLTAHENDLAIAAINTPTSLVISGTPHTVQHITTLCQQQGIKTKTLPTNHA
                    FHSPHTNPILNQLHQHTQTLTYHPPHTPLITANTPPDQLLTPHYWTQQARNTVDYATT
                    TQTLHQHGVTTYIELGPDNTLTTLTHDNLPNTPTTTLTLTHPHHHPQTHLLTNLAKTT
                    TTWHPHHYTHHHNQPHTHTHLDLPTYPFQHHHYWLQPPGKPSDPSPSEGREQATTPST
                    PLRDVLVGKSPQERDEELLRLVRTHAAAVLGHATPEVIVPNKAFKELGFDSLAAIQLR
                    NRLLADVDLPLPATLIFDYPTPMALCQFLRAAIVGADTGTTTRLPLTAVPADEPIAIV
                    GMACRYPGDVRTVDDLWQVVSGGHDAIGGFPTNRGWDLDTLYNPDPDHHGTSYTRSGG
                    FLYDAGNFDPDFFGISPREALAMDPQQRLLLETAWESIEHACINPDSLRGTPTGVFAG
                    LTYHDYAARFPTAPAGFEGYLGHGSAGSIASGRVAYALGLEGPALTVDTACSSSLVAL
                    HLACQALRSGECSMALAGGVTVMSTPAGFVEFSRQRGLAVDGRCKAFSAAADGTGWGE
                    GVGMLLVERLSDARRLGHRILAVVRGSAVNQDGASNGLTAPNGPSQERVIRLALANAD
                    LTPADVDAVEAHGTGTTLGDPIEAQALLATYGQDRPGNEPLWLGSMKSNIGHAQAAAG
                    VGGVIKMVMALRNGLLPRTLHVDEPSPHVDWSAGAVQLLTETVPWPGGEGRLRRAGVS
                    SFGVSGTNAHVILEEAPAHNIPSDTPADDAPGEAAADDVPGEAAGDDAGTGGEATGPA
                    AGSPGVWPWLVSAKSQPALRAQAQALHAHLTDHPGLDLADVGYTLAHARAVFDHRATL
                    IAADRDTFLQALQALAAGEPHPAVIHSSAPGGTGTGEAAGKTAFICSGQGTQRPGMAH
                    GLYHTHPVFAAALNDICTHLDPHLDHPLLPLLTQDPNTQDTTTLEEAAALLQQTPYAQ
                    PALFAFQVALHRLLTDGYHITPHYYAGHSLGEITAAHLAGILTLTDATTLITQRATLM
                    QTMPPGTMTTLHTTPHHITHHLTAHENDLAIAAINTPTSLVISGTPHTVQHITTLCQQ
```

```
QGIKTKTLPTKNAFHSPHTNPILNQLHQHTQTLTYHPPHTPLITANTPPDQLLTPHYW
TQQARNTVDYATTTQTLHQHGVTTYIELGPDNTLTTLTHHNLPNTPTTTLTLTHPHHH
PQTHLLTNLAKTTTTWHPHHYTHHHNQPHTHTHLDLPTYPFQHQHYWLESTQPGAGSG
SGSGSGRAGTAGGTAEVESRFWDAVARQDLETVATTLAVPPSAGLDTVVPALSAWHRH
QHDQARINTWTYQETWKPLTLPTTHQPHQTWLIAIPETQTHHPHITNILTNLHHHGIT
PIPLTLNHTHTNPQHLHHTRQQAQNHTTGPITGLLSLLALDETPHPHHPHTPTGTLLN
LTLTQTHTQTHPPTPLWYATTNATTTHPNDPLTHPTQAQTWGLARTTLLEHPTHTAGI
IDLPTTPTPHTLHHLTQTLTQPHHQTQLAIRTTGTHTRRLTPTTLTPTHQPPTPTPHG
TTLITGGTGALATHLTHHLTTHQPTQHLLLTSRTGPHTPHAQHLTTQLQQKGIHLTIT
TCDTSNPDQLQQLLNTIPPQHPLTTVIHTAGILDDATLTNLTPTQLNNVLRAKAHSAH
LLHQLTQHTPLNAFVLYSSAAATFGAPGQANYAAANAYLDALAHHRHTHHLPATSIAW
GTWQGNGLATGQVSEHLRRRGMFAMPPELAVTAVDGAIASGRPSLLVADIDWKKLGPV
LSSKSSVLLEDLPQAQGTEEARSTVEQTESTNLRQLLMGRSRSEQEEELLSLVRIHSA
AVLGRDDSEAIPPGRLFRDLGFDSLAAVELRNHLAAQTELALPTTLVFDYPSPTKLAQ
FLLSEIAEFQPDNSTPLPRPRAELDEPIAIVGMACRFPGGVTSADDFWDLISSEQDAI
GGFPTDRGWDLDTLYDPDPDHPGTCYTRNGGFLYDAGHFDAEFFGISPREALAMDPQQ
RLLLETAWETIEHAGINPHTLHGTPTGVFTGTNGQDHAAHIRQAPSGTEGFVLTGAAT
SIASGRISYILGLEGPAVTLDTACSSSLVALHLACQSLRSGECTMALAGGATVMTTPI
TFTEFARQRGLAPDGRCKAFSAAADGTGWGEGVGMLLVERLSDARRNGHRVLAVVRGS
AVNQDGASNGLTAPNGPSQQRVIRQALANADLTPADVDAVEAHGTGTTLGDPIEAQAI
LATYGQDRPGNGPLWLGSVKSNVGHTQAAAGVAGVIKMVMALRHRTLPPTLHADEPSP
HVDWSAGAVQLLTETVPWPGGEGRPRRAGVSSFGVSGTNAHVILEEAPADDVPGGPPA
DEDAGSGEEAAAGSPGVWPWLVSAKSQPALRAQAQALHAHLTDHPGLDLADVGYTLAH
ARAVFDHRATLIAADRDTFLQALQALAAGEPHPAVIHSSAPGGTGTGEAAGKTAFICS
GQGTQRPGMAHGLYHTHPVFAAALNDICTHLDPHLDHPLLPLLTQNDNDNDNEDAAAL
LQQTPYAQPALFAFQVALHRLLTDGYHITPHYYAGHSLGEITAAHLAGILTLTDATTL
ITQRATLMQTMPPGTMTTLHTTPHHITHHLTAHENDLAIAAINTPTSLVISGTPHTVQ
HITTLCQQQGIKTKTLPTNHAFHSPHTNPILNQLHQHTQTLTYHPPHTPLITANTPPD
QLLTPHYWTQQARNTVDYATTTQTLHQHGVTTYIELGPDNTLTTLTHHNLPNTPTTTL
TLTHPHHHPQTHLLTNLAKTTTTWHPHHYTHHHNQPHTHTHLDLPTYPFQHHHYWLEL
PSAQTSPGQRRSRRSAPDTAESEFWDAVNEEDLQSLAETLDIDASALDTVVPALSAWH
RHQHDQARINTWTYQETWKPLTLPTTHQPHQTWLIAIPETQTHHPHITNILTNLHHHG
ITPIPLTVNHTHTNPQHLHHTLHHTRQQAQNHTTGPITGLLSLLALDETPHPHHPHTP
TGTLLNLTLPQTHTQTHPPTPLWYATTNATTTHPNDPLTHPTQAQTWGLARTTLLEHP
THTAGIIDLPTTPTPHTLHHLTQTLTQPHHQTQLAIRTTGTHTRRLTPTTLTPTHQPP
TPTPHGTTLITGGTGALATHLTHHLTTHQPTQHLLLTSRTGPHTPHAQHLTTQLQQKG
IHLTITTCDTSNPDQLQQLLNTIPPQHPLTTVIHTAGVNLFAPVSETDAESFSSVTAA
KATGAAILHELLLDHETLEHFILFSSGAGAWGSGNQCAYSAANAYLDALATHRQTHGL
PGASIAWGPWAGKGMSAGDAAHGYLEKRGILPMEPRMALAAFHRARAQRPNSNLIIAD
IDWERFVPAFTARRHSPLIEDIPEVRQAAQELEAAASTAKTTTAQPIATSLRERLARL
TSSKQNQVLLGLIRTGICTVLGLRNPEGIEDQRAFRDLGFDSLTSAQFSKELAKETGL
PLPPSLVFDYPTPQECAAHLRTQLVDLDDEEDAALSNALPQVAHRRTVEDEPIAIIGM
ACRFPGGVRSADDLWELLASGKDAIGVFPTDRGWDLDTLYDPDPDHPGTCYTRNGGFL
YGAGHFDAEFFGISPREALAMDPQQRLLLETAWETIEHAGINPHTLHGTPTGVFAGIN
AQDHAAHIRQSRDVETIEGYALTGSSGSVASGRVAYTLGLEGPAVSVDTACSSSLVAL
```

Query Match 100.0%; Score 1290; DB 1; Length 64957;
Best Local Similarity 100.0%; Pred. No. 6.6e-151;
Matches 1290; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 GAGCCGATCGCCATCGTTGGCATGGCCTGTCGTTTCCCCGGCGGAGTGACCTCGGCGGAC 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    6356 GAGCCGATCGCCATCGTTGGCATGGCCTGTCGTTTCCCCGGCGGAGTGACCTCGGCGGAC 6415

Qy      61 GACTTCTGGGATCTGATCTCCTCCGAGCAGGACGCGATCGGCGGATTCCCCACCGACCGC 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    6416 GACTTCTGGGATCTGATCTCCTCCGAGCAGGACGCGATCGGCGGATTCCCCACCGACCGC 6475

Qy     121 GGCTGGGACCTGGACACGCTCTACGACCCCGACCCCGACCACCCCGGCACCTGCTACACC 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    6476 GGCTGGGACCTGGACACGCTCTACGACCCCGACCCCGACCACCCCGGCACCTGCTACACC 6535

Qy     181 CGAAACGGCGGATTCCTCTACGACGCAGGCCACTTCGACGCCGAATTCTTCGGCATCAGC 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    6536 CGAAACGGCGGATTCCTCTACGACGCAGGCCACTTCGACGCCGAATTCTTCGGCATCAGC 6595

Qy     241 CCCCGCGAAGCCCTCGCCATGGACCCCCAGCAACGACTCCTCCTCGAAACCGCCTGGGAA 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    6596 CCCCGCGAAGCCCTCGCCATGGACCCCCAGCAACGACTCCTCCTCGAAACCGCCTGGGAA 6655

Qy     301 ACCATCGAACACGCCGGCATCAACCCCCACACCCTCCACGGCACCCCCACCGGAGTCTTC 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    6656 ACCATCGAACACGCCGGCATCAACCCCCACACCCTCCACGGCACCCCCACCGGAGTCTTC 6715

Qy     361 ACCGGCACCAACGGACAGGACTACGCACTTCGCGTGCACAACGCGGGCCAGTCAACCGAT 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    6716 ACCGGCACCAACGGACAGGACTACGCACTTCGCGTGCACAACGCGGGCCAGTCAACCGAT 6775

Qy     421 GGTTTCGCACTGACCGGAACCGCCGGCAGCGTCATCTCCGGTCGTATCTCGTACACGTTT 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    6776 GGTTTCGCACTGACCGGAACCGCCGGCAGCGTCATCTCCGGTCGTATCTCGTACACGTTT 6835

Qy     481 GGTTTTGAGGGTCCTGCGGTGTCGGTGGACACGGCTTGTTCCTCGTCGTTGGTGGCTTTG 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    6836 GGTTTTGAGGGTCCTGCGGTGTCGGTGGACACGGCTTGTTCCTCGTCGTTGGTGGCTTTG 6895

Qy     541 CATCTGGCCTGTCAGGCGTTGCGTGCGGGTGAGTGCTCGATGGCGCTTGCCGGGGGTGTG 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    6896 CATCTGGCCTGTCAGGCGTTGCGTGCGGGTGAGTGCTCGATGGCGCTTGCCGGGGGTGTG 6955

Qy     601 ACGGTGATGTCGTCTCCGGGTGCCTTCGTGGAGTTTTCGCGGCAGCGGGGTCTGGCCGCG 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    6956 ACGGTGATGTCGTCTCCGGGTGCCTTCGTGGAGTTTTCGCGGCAGCGGGGTCTGGCCGCG 7015

Qy     661 GACGGGCATTGCAAGGCGTTCTCGGCGGCGGCGGACGGGACCGGCTGGGGTGAGGGTGTG 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    7016 GACGGGCATTGCAAGGCGTTCTCGGCGGCGGCGGACGGGACCGGCTGGGGTGAGGGTGTG 7075

Qy     721 GGGATGCTGCTGGTGGAGCGGCTCTCCGACGCCCATCGCAACGGTCACCGTGTCCTGGCC 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    7076 GGGATGCTGCTGGTGGAGCGGCTCTCCGACGCCCATCGCAACGGTCACCGTGTCCTGGCC 7135

Qy     781 GTGGTGCGTGGCAGTGCGGTCAACCAGGACGGTGCGAGCAACGGTCTGACCGCGCCCAAC 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    7136 GTGGTGCGTGGCAGTGCGGTCAACCAGGACGGTGCGAGCAACGGTCTGACCGCGCCCAAC 7195

Qy     841 GGGCCGTCCCAGCAGCGTGTCATCCGCCAGGCCCTCGCCAACGCCGGCTTGTCGGCCGGT 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    7196 GGGCCGTCCCAGCAGCGTGTCATCCGCCAGGCCCTCGCCAACGCCGGCTTGTCGGCCGGT 7255

Qy     901 GATGTCGACGCGGTGGAGGCCCACGGCACCGGCACCACTTTGGGCGACCCGATCGAGGCC 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    7256 GATGTCGACGCGGTGGAGGCCCACGGCACCGGCACCACTTTGGGCGACCCGATCGAGGCC 7315

Qy     961 CAGGCCCTCCTCGCGACCTACGGACAGGACCGTGCCGGCGAGGGGCCGCTGTGGCTGGGC 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    7316 CAGGCCCTCCTCGCGACCTACGGACAGGACCGTGCCGGCGAGGGGCCGCTGTGGCTGGGC 7375

Qy    1021 TCGGTCAAGTCCAATGTCGGTCACACACAGGCTGCCGCGGGCGTCGCCGGGGTGATCAAG 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    7376 TCGGTCAAGTCCAATGTCGGTCACACACAGGCTGCCGCGGGCGTCGCCGGGGTGATCAAG 7435

Qy    1081 ATGGTGATGGCGCTGCGGCATGGTCTGCTGCCGCGGACGTTGCATGTGGATGAGCCGTCG 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    7436 ATGGTGATGGCGCTGCGGCATGGTCTGCTGCCGCGGACGTTGCATGTGGATGAGCCGTCG 7495

Qy    1141 CCGCATGTGGACTGGTCCGCGGGTGCGGTGCAGCTGCTGACGGAGACGGTGCCCTGGCCC 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    7496 CCGCATGTGGACTGGTCCGCGGGTGCGGTGCAGCTGCTGACGGAGACGGTGCCCTGGCCC 7555

Qy    1201 GGCGGGGAGGGGCGGCTACGGCGGGCAGGAGTGTCATCATTCGGCGTCAGCGGCACCAAC 1260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    7556 GGCGGGGAGGGGCGGCTACGGCGGGCAGGAGTGTCATCATTCGGCGTCAGCGGCACCAAC 7615

Qy    1261 GCCCACGTCATCCTCGAAGAAGCACCCGCC 1290
          ||||||||||||||||||||||||||||||
Db    7616 GCCCACGTCATCCTCGAAGAAGCACCCGCC 7645
```

RESULT 5
E38021
LOCUS E38021 31422 bp DNA linear PAT 31-JAN-2002
DEFINITION Avermectin aglycon synthase gene.
ACCESSION E38021
VERSION E38021.1 GI:18626910
KEYWORDS JP 2000245457-A/2.

```
SOURCE      Streptomyces avermitilis.
  ORGANISM  Streptomyces avermitilis
            Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
            Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1  (bases 1 to 31422)
  AUTHORS   Omura,S. and Ikeda,H.
  TITLE     Avermectin aglycon synthase gene
  JOURNAL   Patent: JP 2000245457-A 2 12-SEP-2000;
            THE KITASATO INSTITUTE
COMMENT     OS   Streptomyces avermitilis
            PN   JP 2000245457-A/2
            PD   12-SEP-2000
            PF   24-FEB-1999 JP 1999046961
            PR
            PI   SATOSHI OMURA,HARUO IKEDA
            PC   C12N15/00,A61K31/70,C12N1/15,C12N9/88,C12P19/62, PC
            C12Q1/68//C07D493/22,
            PC   (C12N1/15,C12R1:465),(C12N9/88,C12R1:465),C12N15/00 CC
            FH   Key             Location/Qualifiers
            FT   CDS             (1). .(14643)
            FT   CDS             (14824). .(31419).
FEATURES             Location/Qualifiers
     source          1. .31422
                     /organism="Streptomyces avermitilis"
                     /db_xref="taxon:33903"
BASE COUNT     4136 a  10237 c  11677 g   5372 t
ORIGIN

  Query Match           80.0%;  Score 1032.2;  DB 6;  Length 31422;
  Best Local Similarity   87.9%;  Pred. No. 5.2e-119;
  Matches 1137;  Conservative    0;  Mismatches  153;  Indels    3;  Gaps    1;

Qy          1 GAGCCGATCGCCATCGTTGGCATGGCCTGTCGTTTCCCCGGCGGAGTGACCTCGGCGGAC 60
              ||||| |||| ||| |||||||||| || || |  ||||| || || ||||| || ||
Db      20413 GAGCCCGTCGCGATCATTGGCATGGCGTGCCGCTATCCCGGAGGCGTCACCTCAGCCGAG 20472

Qy         61 GACTTCTGGGATCTGATCTCCTCCGAGCAGGACGCGATCGGCGGATTCCCCACCGACCGC 120
              ||  | ||||| ||| || | || | |  |||| || ||||||  || || || |||||
Db      20473 GAGCTGTGGGAACTGCTCGCATCGGGGAGGGACACGGTCGGCGAGTTTCCGACGGACCGT 20532

Qy        121 GGCTGGGACCTGGACACGCTCTACGACCCCGACCCCGACCACCCCGGCACCTGCTACACC 180
              || ||||||||||||  | || | ||| || || || |  |  || |||||||  ||||||
Db      20533 GGGTGGGACCTGGAAGCACTGTTCGATCCGGAACCGGGTCGGCCGGGCACCTCGTACACC 20592

Qy        181 CGAAACGGCGGATTCCTCTACGACGCAGGCCACTTCGACGCCGAATTCTTCGGCATCAGC 240
              ||    ||  | ||||||||||||||| ||  | ||||||||||  |||||||| |||||
Db      20593 CGCTGTGGGAGTTTCCTCTACGACGCGGGGGAGTTCGACGCCGGCTTCTTCGGGATCAGT 20652

Qy        241 CCCCGCGAAGCCCTCGCCATGGACCCCCAGCAACGACTCCTCCTCGAAACCGCCTGGGAA 300
              || || || || || || |||||||| ||||| ||| | || || ||  || | |||||
Db      20653 CCGCGTGAGGCACTGGCGATGGACCCGCAGCAGCGATTGCTGCTGGAGGCCTCATGGGAG 20712

Qy        301 ACCATCGAACACGCCGGCATCAACCCCCACACCCTCCACGGCACCCCCACCGGAGTCTTC 360
              ||||  || || || || ||  ||||     ||| | | ||| | ||  || || || |||
Db      20713 GCCATGGAGCAGGCAGGTATTGACCCTACGACCGTACGCGGGAGCCAGACAGGCGTGTTC 20772

Qy        361 ACCGGCACCAACGGACAGGACTACGCACTTCGCGTGCAC---AACGCGGGCCAGTCAACC 417
              | |||  ||     |||| ||| | ||   |  |||||   ||||| |        |||
Db      20773 GCGGGCCTCATTCCGCAGGCCTATGGACCCAGGCTGCACGAAAACGCCGCAGCCGACACC 20832

Qy        418 GATGGTTTCGCACTGACCGGAACCGCCGGCAGCGTCATCTCCGGTCGTATCTCGTACACG 477
              || || |  | |||||||| ||  |||| || ||  ||||||||||||||||||||||
Db      20833 GAGGGCTATGTCCTGACCGGCACATCCGGGAGTGTGGCCTCCGGTCGTATCTCGTACACG 20892

Qy        478 TTTGGTTTTGAGGGTCCTGCGGTGTCGGTGGACACGGCTTGTTCCTCGTCGTTGGTGGCT 537
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      20893 TTTGGTTTTGAGGGTCCTGCGGTGTCGGTGGACACGGCTTGTTCCTCGTCGTTGGTGGCT 20952

Qy        538 TTGCATCTGGCCTGTCAGGCGTTGCGTGCGGGTGAGTGCTCGATGGCGCTTGCCGGGGGT 597
              || ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      20953 TTACATCTGGCCTGTCAGGCGTTGCGTGCGGGTGAGTGCTCGATGGCGCTTGCCGGGGGT 21012

Qy        598 GTGACGGTGATGTCGTCTCCGGGTGCCTTCGTGGAGTTTTCGCGGCAGCGGGGTCTGGCC 657
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      21013 GTGACGGTGATGTCGTCTCCGGGTGCCTTCGTGGAGTTTTCGCGGCAGCGGGGTCTGGCC 21072

Qy        658 GCGGACGGGCATTGCAAGGCGTTCTCGGCGGCGGCGGACGGGACCGGCTGGGGTGAGGGT 717
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      21073 GCGGACGGGCATTGCAAGGCGTTCTCGGCGGCGGCGGACGGGACCGGCTGGGGTGAGGGT 21132

Qy        718 GTGGGGATGCTGCTGGTGGAGCGGCTCTCCGACGCCCATCGCAACGGTCACCGTGTCCTG 777
              |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db      21133 GTGGGGATGCTGCTGGTGGAGCGGCTCTCCGACGCCCGTCGCAACGGTCACCGTGTCCTG 21192

Qy        778 GCCGTGGTGCGTGGCAGTGCGGTCAACCAGGACGGTGCGAGCAACGGTCTGACCGCGCCC 837
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      21193 GCCGTGGTGCGTGGCAGTGCGGTCAACCAGGACGGTGCGAGCAACGGGCTGACCGCGCCC 21252

Qy        838 AACGGGCCGTCCCAGCAGCGTGTCATCCGCCAGGCCCTCGCCAACGCCGGCTTGTCGGCC 897
              |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      21253 AACGGGCCCTCCCAGCAGCGTGTCATCCGCCAGGCCCTCGCCAACGCCGGCTTGTCGGCC 21312

Qy        898 GGTGATGTCGACGCGGTGGAGGCCCACGGCACCGGCACCACTTTGGGCGACCCGATCGAG 957
              |||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db      21313 GGTGATGTCGATGCGGTGGAGGCCCACGGCACCGGCACCACTTTGGGCGACCCGATCGAG 21372

Qy        958 GCCCAGGCCCTCCTCGCGACCTACGGACAGGACCGTGCCGGCGAGGGGCCGCTGTGGCTG 1017
              |||||||||||||| |||||||||| ||||||||||||||||||||||||||||||||||
Db      21373 GCCCAGGCCCTCCTTGCGACCTACGGGCAGGACCGTGCCGGCGAGGGGCCGCTGTGGCTG 21432

Qy       1018 GGCTCGGTCAAGTCCAATGTCGGTCACACACAGGCTGCCGCGGGCGTCGCCGGGGTGATC 1077
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      21433 GGCTCGGTCAAGTCCAATGTCGGTCACACACAGGCTGCCGCGGGCGTCGCCGGGGTGATC 21492

Qy       1078 AAGATGGTGATGGCGCTGCGGCATGGTCTGCTGCCGCGGACGTTGCATGTGGATGAGCCG 1137
              |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db      21493 AAGATGGTGATGGCGCTGCGGAATGGTCTGCTGCCGCGGACGTTGCATGTGGATGAGCCG 21552

Qy       1138 TCGCCGCATGTGGACTGGTCCGCGGGTGCGGTGCAGCTGCTGACGGAGACGGTGCCCTGG 1197
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      21553 TCGCCGCATGTGGACTGGTCCGCGGGTGCGGTGCAGCTGCTGACGGAGACGGTGCCCTGG 21612

Qy       1198 CCCGGCGGGGAGGGGCGGCTACGGCGGGCAGGAGTGTCATCATTCGGCGTCAGCGGCACC 1257
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      21613 CCCGGCGGGGAGGGGCGGCTACGGCGGGCAGGAGTGTCATCATTCGGCGTCAGCGGCACC 21672

Qy       1258 AACGCCCACGTCATCCTCGAAGAAGCACCCGCC 1290
              ||||||||||||||||||||||||||||||||
Db      21673 AACGCCCACGTCATCCTCGAAGAAGCACCCGCC 21705


RESULT 6
AB032367/c
LOCUS       AB032367               64957 bp    DNA     linear   BCT 18-SEP-1999
DEFINITION  Streptomyces avermitilis polyketide synthase gene cluster (aveA1,
            aveA2, aveA3, aveA4) and aveC, aveE genes, complete cds.
ACCESSION   AB032367
VERSION     AB032367.1  GI:5902890
KEYWORDS    AveC; cytochrome P450 hydroxylase; type I polyketide synthase AVES
            4; type I polyketide synthase AVES 3; type I polyketide synthase
            AVES 2; type I polyketide synthase AVES 1.
SOURCE      Streptomyces avermitilis DNA.
  ORGANISM  Streptomyces avermitilis
            Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
            Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1  (bases 1 to 64957)
  AUTHORS   Ikeda,H., Nonomiya,T., Usami,M., Ohta,T. and Omura,S.
  TITLE     Organization of the biosynthetic gene cluster for the polyketide
            anthelmintic macrolide avermectin in Streptomyces avermitilis
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9509-9514 (1999)
  MEDLINE   99380548
REFERENCE   2  (bases 1 to 64957)
  AUTHORS   Ikeda,H., Nonomiya,T., Usami,M., Ohta,T. and Omura,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-SEP-1999) Haruo Ikeda, School of Pharmaceutical
            Sciences, Kitasato University, Microbial chemistry; 5-9-1
```

Shirokane, Minato-ku, Tokyo 108-8641, Japan
(E-mail:ikeda@mc.pharm.kitasato-u.ac.jp, Tel:+81-3-5791-6242,
Fax:+81-3-3444-6197)

FEATURES Location/Qualifiers

source 1. .64957
/organism="Streptomyces avermitilis"
/db_xref="taxon:33903"

gene 101. .12019
/gene="aveA1"

CDS 101. .12019
/gene="aveA1"
/note="contains a loading module, Ave module 1 and module 2
multifunctional polyketide synthase"
/codon_start=1
/transl_table=11
/product="type I polyketide synthase AVES 1"
/protein_id="BAA84474.1"
/db_xref="GI:5902891"
/translation="MQRMDGGEEPRPAAGEVLGVADEADGGVVFVFPGQGPQWPGMGR
ELLDASDVFRESVRACEAAFAPYVDWSVEQVLRDSPDAPGLDRVDVVQPTLFAVMISL
AALWRSQGVEPCAVLGHSLGEIAAAHVSGGLSLADAARVVTLWSQAQTTLAGTGALVS
VAATPDELLPRIAPWTEDNPARLAVAAVNGPRSTVVSGAREAVADLVADLTAAQVRTR
MIPVDVPAHSPLMYAIEERVVSGLLPITPRPSRIPFHSSVTGGRLDTRELDAAYWYRN
MSSTVRFEPAARLLLQQGPKTFVEMSPHPVLTMGLQELAPDLGDTTGTADTVIMGTLR
RGQGTLDHFLTSLAQLRGHGETSATTVLSARLTALSPTQQQSLLLDLVRAHTMAVLND
DGNERTASDAGPSASFAHLGFDSVMGVELRNRLSKATGLRLPVTLIFDHTTPAAVAAR
LRTAALGHLDEDTAPVPDSPSGHGGTAAADDPIAIIGMACRFPGGVRSPKDLWELAAS
GGDAIGPFPTDRGWPTEQRHAQDPTQPGTFYPQGGGFLHDAAHFDAGFFGISPREALA
MDPQQRLLLETSWEAFERAGIDPLSVRGSRTGVFAGALSFDYGPRMDTASSEGAADVE
GHILTGTTGSVLSGRIAYSFGLEGPAITVDTGCSASLVTLHLACQSLRSGECTLALAG
GVSVMSTLGMFIEFSRQRGLSVDGRCKAYSAAADGTGWGEGVGMLLVERLSDAVRLGH
RVLAVVRGSAVNQDGASNGLTAPNGPAQERVIRQALANAGLSVADVDVVEGHGTGTTL
GDPIEAQALLATYGQRAGDRPLWLGSLKSNIGHTMAAAGVGGVIKMVMALREGVLPRT
LHVDKPSPQVDWSAGAVRLLTEAVPWPGDAAGRLRRAGVSSFGIGGTNAHVILEEAPA
AGGCVAGGGVLEGAPGLAISVAESVAAPVAVSAPVAESVPVPVPVPVPVPVSARSEAG
LRAQAEALRQYVAVRPDVSLADVGAGLACGRAVLEHRAVVLAADREELVQGLGALAAG
EPDRRVTTGHAPGGDRGGVVFVFPGQGGQWAGMGVRLLASSPVFARRMQACEEALAPW
VDWSVVDILRRDAGDAVWERADVVQPVLFSVMVSLAALWRSYGIEPDAVLGHSQGEIA
AAHVCGALSLKDAAKTVALRSRALAAVRGRGGMASVPLPAQEVEQLIGERWAGRLWVA
AVNGPRSTAVSGDAEAVDEVLAYCAGTGVRARRIPVDYASHCPHVQPLREELLELLGD
ISPQPSGVPFFSTVEGTWLDTTTLDAAYWYRNLHQPVRFSDAVQALADDGHRVFVEVS
PHPTLVPAIEDTTEDTAEDVTAIGSLRRGDNDTRRFLTALAHTHTTGIGTPTTWHHHY
THHHTHPHPHTHLDLPTYPFQHQHYWLESSQPGAGSGSGAGAGSGAGSGRAGTAGGTA
EVESRFWDAVARQDLETVATTLAVPPSAGLDTVVPALSAWHRHQHDQARINTWTYQET
WKPLTLPTTHQPHQTWLIAIPETQTHHPHITNILTNLHHHGITPIPLTLNHTHTNPQH
LHHTLHHTRQQAQNHTTGAITGLLSLLALDETPHPHHPHTPTGTLLNLTLTQTHTQTH
PPTPLWYATTNATTTHPNDPLTHPTQAQTWGLARTTLLEHPTHTAGIIDLPTTPTPHT
LQHLTQTLTQPHHQTQLAIRTTGTHTRRLTPTTLTPTHQPPTPTPHTGTTLITGGTGAL
ATHLTHHLTTHQPTQHLLLTSRTGPHTPHAQHLTTQLQQKGIHLTITTCDTSNPDQLQ
QLLNTIPPQHPLTTVIHTAGILDDATLTNLTPTQLNNVLRAKAHSAHLLHQLTQHTPL
TAFVLYSSAAATFGAPGQANYAAANAYLDALAHHRHTHHLPATSIAWGTWQGNGLADS
DKARAYLDRRGFRPMSPELATAAVTQAIADTERPYVVIADIDWSKIEHTSQTSDLVSA
ARELREPAVQRPTPPAELHKTLAHQTSADQRAALLELVRDHVAAVLRHADPKAIAPDQS
FRALGFDSLTAVEFRNLLIKATGLRLPVSLVFDHPTPAKLAVHLQNQLRGTAAESAPS
AAAVTAEASVTEPIAIVGMACRFPGGVTSADDFWDLISSEQDAIGGFPTDRGWDLDTL
YDPDPDHPGTCYTRNGGFLYDAGHFDAEFFGISPREALAMDPQQRLLLETAWETIEHA
GINPHTLHGTPTGVFTGTNGQDYALRVHNAGQSTDGFALTGTAGSVISGRISYTFGFE
GPAVSVDTACSSSLVALHLACQALRAGECSMALAGGVTVMSSPGAFVEFSRQRGLAAD
GHCKAFSAAADGTGWGEGVGMLLVERLSDAHRNGHRVLAVVRGSAVNQDGASNGLTAP
NGPSQQRVIRQALANAGLSAGDVDAVEAHGTGTTLGDPIEAQALLATYGQDRAGEGPL
WLGSVKSNVGHTQAAAGVAGVIKMVMALRHGLLPRTLHVDEPSPHVDWSAGAVQLLTE
TVPWPGGEGRLRRAGVSSFGVSGTNAHVILEEAPADDVPGGPPAGEGDAGSDDEAAAG
SPGVWPWLVSAKSQPALRAQAQALHAHLTDHPGLDLADVGYTLAHARAVFDHRATLIA
ADRDTFLQALQALAAGEPHPAVIHSSAPGGTGTGEAAGKTAFICSGQGTQRPGMAHGL
YHTHPVFAAALNDICTHLDPHLDHPLLPLLTQNDNDNEDAAALLQQTRYAQPALFAFQ
VALHRLLTDGYHITPHYYAGHSLGEITAAHLAGILTLTDATTLITQRATLMQTMPPGT
MTTLHTTPHHITHHLTAHENDLAIAAINTPTSLVISGTPHTVQHITTLCQQQGIKTKT
LPTNHAFHSPHTNPILNQLHQHTQTLTYHPPHTPLITANTPPDQLLTPHYWTQQARNT
VDYATTTQTLHQHGVTTYIELGPDNTLTTLTHHNLPNPPTTTLTLTHPHHHPQTHLLT
NLAKTTTTWHPHHYTHHDNQPHTHTHLDLPTYPFQHHHYWLESTQPGAGNVSAAGLDP
TEHPLLGATLELATDGGALLAGRLSLRSHPWLADHAVGGTVLLSGATFLELALHAGTY
VGCDRVDELTLHAPLVVPVDGGVSVQVGVAAADGEGRRLVSVYARGGSACGGGGASGG
VWTCHASGVLVEAAAGGVVVDGLAGVWPPRGAVAVDVDGVRDRLAGAGCVLGPVFSGL
RAVWRDGGDLLAEVCLPEEAWGDAAGFGLHPALLDGVVQPLSVLLPGGTGFGEGAGFG
EGVRVPAVWGGVSLHRAGVTGVRVRVSAVGRGGGREAVSVVVGDEAGVPVASVDRLEL
RPVDMGQLRAVSVSAGRRGSLYAVQWAEVGPVPVCGQAWAWHEDVGESGGGPVPGVVV
LRCPDAGAGGGGGGGGGGVGEVVGGVLGVVQGWLGLERFAGSRLVVVTRGAVVAGPE
DGPVDVVGASVWGLVRSAQAEHPDRFVLLDLDTDTGTDLDTGAGAGWGVDGGRVAAVV
ACGEPQLAVRGERLLAARLKRLESSGDVPAQRSGDTRARRSDVPAQRSGGVPARRSVD
VSGREVLPWLSGGSVLVTGGTGVLGAAVARHLAGVCGVRDLLLVSRRGPDAPGAEGLR
AELAALGAEVRIVACDVGERREVVRLLEGVPAGCPLTGVVHAAGVLDDATIASLTPER
LGTVFAAKVDAALLLDELTRGMELSAFVLFSSAAGILGSAGQGNYAAANAALDALAYR
RRAAGLPGVSLAWGLWEEASGMTGHLAGTDHRRIIRSGLHPMSTPDALALFDAALALD
RPVLLPADLRPAPPLPPLLQDLLPATRRRTTRTTTTGGADNGAQLHARLAGQTHEQQH
TTLLALVRSHIATVLGHTTPDTIPPDRAFRDLGFDSLTAVELRNRLSRTTGLRLPTTL
AFDHPNPTTLTHHLHTQLQPQPDNAVAPVLAELDKLESALSALDKTDSASERVTLRLK
SLMLRWNAPQHPTAESADDDEKFTSATEAEIFKFIDNDLGLS"

gene 12071. .30790
/gene="aveA2"

CDS 12071. .30790
/gene="aveA2"
/note="contains Ave module 3, module 4, module 5 and module 6
multifunctional polyketide synthase"
/codon_start=1
/transl_table=11
/product="type I polyketide synthase AVES 2"
/protein_id="BAA84475.1"
/db_xref="GI:5902892"
/translation="MQLANEAKLLEYLKRVTADLDRTRRRLYEVVEREQEPIAIVGMA
CRYPGGATSPTRLWHLVKSQTDAIGEFPTDRGWNLEQLYDPDPDRSGTSYTRSGGFLY
DAGDFDAAFFELSPREALAMDPQQRLLLETTWETFEQGGIDPRSMRGSRTGVFVGINP
EDYTTGYTHQPSNAVEGYLLTGSAASIASGRISYNFGLEGPAITIDTACSSSLVALHL
ACQALRSGECTMALAGGASVMATPFVFTEFSRQRGLAADGRCKAFSAAADGTGWSEGV
GMLLVERLSDARRNGHRVLAVVRGSAVNQDGASNGLTAPNGRSQVKVIRQALANAHLS
PADVDAVEAHGTGTTLGDPIEAQALVEAYGQDRPNGRPLWLGTLKSNIGHSMAAAGVG
GVIKMVMALRNGLLPRTLHVDEPSPHVDWSAGAVQLLTETVPWPGGEGRLRRAGVSSF
GVSGTNAHVILEEAPAHNIPSDTPADDVPGESAADEDAGSGDEAAAGSPGVWPWLVSA
KSQPALRAQAQALHAHLTDHPGLDLADVGYTLAHARAVFDHRATLIAADRDTFLQALQ
ALAAGEPHPAVIHSSAPGGTGTGEAAGKTAFICSGQGTQRPGMAHGLYHTHPVFAAAL
NDICTHLDPHLDHPLLPLLTQDPNTQDTTTLEEAAALLQQTRYAQPALFAFQVALHRL
LTDGYHITPHYYAGHSLGEITAAHLAGILTLTDATTLITQRATLMQTMPPGTMTTLHT
TPHHITHHLTAHENDLAIAAINTPTSLVISGTPHTVQHITTLCQQQGIKTKTLPTNHA
FHSPHTNPILNQLHQHTQTLTYHPPHTPLITANTPPDQLLTPHYWTQQARNTVDYATT
TQTLHQHGVTTYIELGPDNTLTTLTHDNLPNTPTTTLTLTHPHHHPQTHLLTNLAKTT
TTWHPHHYTHHHNQPHTHTHLDLPTYPFQHHHYWLQPPGKPSDPSPSEGREQATTPST
PLRDVLVGKSPQERDEELLRLVRTHAAAVLGHATPEVIVPNKAFKELGFDSLAAIQLR
NRLLADVDLPLPATLIFDYPTPMALCQFLRAAIVGADTGTTTRLPLTAVPADEPIAIV
GMACRYPGDVRTVDDLWQVVSGGHDAIGGFPTNRGWDLDTLYNPDPDHHGTSYTRSGG
FLYDAGNFDPDFFGISPREALAMDPQQRLLLETAWESIEHACINPDSLRGTPTGVFAG
LTYHDYAARFPTAPAGFEGYLGHGSAGSIASGRVAYALGLEGPALTVDTACSSSLVAL
HLACQALRSGECSMALAGGVTVMSTPAGFVEFSRQRGLAVDGRCKAFSAAADGTGWGE
GVGMLLVERLSDARRLGHRILAVVRGSAVNQDGASNGLTAPNGPSQERVIRLALANAD
LTPADVDAVEAHGTGTTLGDPIEAQALLATYGQDRPGNEPLWLGSMKSNIGHAQAAAG
VGGVIKMVMALRNGLLPRTLHVDEPSPHVDWSAGAVQLLTETVPWPGGEGRLRRAGVS
SFGVSGTNAHVILEEAPAHNIPSDTPADDAPGEAAADDVPGEAAGDDAGTGGEATGPA
AGSPGVWPWLVSAKSQPALRAQAQALHAHLTDHPGLDLADVGYTLAHARAVFDHRATL
IAADRDTFLQALQALAAGEPHPAVIHSSAPGGTGTGEAAGKTAFICSGQGTQRPGMAH
GLYHTHPVFAAALNDICTHLDPHLDHPLLPLLTQDPNTQDTTTLEEAAALLQQTPYAQ
PALFAFQVALHRLLTDGYHITPHYYAGHSLGEITAAHLAGILTLTDATTLITQRATLM
QTMPPGTMTTLHTTPHHITHHLTAHENDLAIAAINTPTSLVISGTPHTVQHITTLCQQ
QGIKTKTLPTKNAFHSPHTNPILNQLHQHTQTLTYHPPHTPLITANTPPDQLLTPHYW
TQQARNTVDYATTTQTLHQHGVTTYIELGPDNTLTTLTHHNLPNTPTTTLTLTHPHHH
PQTHLLTNLAKTTTTWHPHHYTHHHNQPHTHTHLDLPTYPFQHQHYWLESTQPGAGSG
SGSGSGRAGTAGGTAEVESRFWDAVARQDLETVATTLAVPPSAGLDTVVPALSAWHRH
QHDQARINTWTYQETWKPLTLPTTHQPHQTWLIAIPETQTHHPHITNILTNLHHHGIT
PIPLTLNHTHTNPQHLHHTRQQAQNHTTGPITGLLSLLALDETPHPHHPHTPTGTLLN
LTLTQTHTQTHPPTPLWYATTNATTTHPNDPLTHPTQAQTWGLARTTLLEHPTHTAGI
IDLPTTPTPHTLHHLTQTLTQPHHQTQLAIRTTGTHTRRLTPTTLTPTHQPPTPTPHG
TTLITGGTGALATHLTHHLTTHQPTQHLLLTSRTGPHTPHAQHLTTQLQQKGIHLTIT
TCDTSNPDQLQQLLNTIPPQHPLTTVIHTAGILDDATLTNLTPTQLNNVLRAKAHSAH
LLHQLTQHTPLNAFVLYSSAAATFGAPGQANYAAANAYLDALAHHRHTHHLPATSIAW
GTWQGNGLATGQVSEHLRRRGMFAMPPELAVTAVDGAIASGRPSLLVADIDWKKLGPV
LSSKSSVLLEDLPQAQGTEEARSTVEQTESTNLRQLLMGRSRSEQEEELLSLVRIHSA
AVLGRDDSEAIPPGRLFRDLGFDSLAAVELRNHLAAQTELALPTTLVFDYPSPTKLAQ

```
                     FLLSEIAEFQPDNSTPLPRPRAELDEPIAIVGMACRFPGGVTSADDFWDLISSEQDAI
                     GGFPTDRGWDLDTLYDPDPDHPGTCYTRNGGFLYDAGHFDAEFFGISPREALAMDPQQ
                     RLLLETAWETIEHAGINPHTLHGTPTGVFTGTNGQDHAAHIRQAPSGTEGFVLTGAAT
                     SIASGRISYILGLEGPAVTLDTACSSSLVALHLACQSLRSGECTMALAGGATVMTTPI
                     TFTEFARQRGLAPDGRCKAFSAAADGTGWGEGVGMLLVERLSDARRNGHRVLAVVRGS
                     AVNQDGASNGLTAPNGPSQQRVIRQALANADLTPADVDAVEAHGTGTTLGDPIEAQAI
                     LATYGQDRPGNGPLWLGSVKSNVGHTQAAAGVAGVIKMVMALRHRTLPPTLHADEPSP
                     HVDWSAGAVQLLTETVPWPGGEGRPRRAGVSSFGVSGTNAHVILEEAPADDVPGGPPA
                     DEDAGSGEEAAAGSPGVWPWLVSAKSQPALRAQAQALHAHLTDHPGLDLADVGYTLAH
                     ARAVFDHRATLIAADRDTFLQALQALAAGEPHPAVIHSSAPGGTGTGEAAGKTAFICS
                     GQGTQRPGMAHGLYHTHPVFAAALNDICTHLDPHLDHPLLPLLTQNDNDNDNEDAAAL
                     LQQTPYAQPALFAFQVALHRLLTDGYHITPHYYAGHSLGEITAAHLAGILTLTDATTL
                     ITQRATLMQTMPPGTMTTLHTTPHHITHHLTAHENDLAIAAINTPTSLVISGTPHTVQ
                     HITTLCQQQGIKTKTLPTNHAFHSPHTNPILNQLHQHTQTLTYHPPHTPLITANTPPD
                     QLLTPHYWTQQARNTVDYATTTQTLHQHGVTTYIELGPDNTLTTLTHHNLPNTPTTTL
                     TLTHPHHHPQTHLLTNLAKTTTTWHPHHYTHHHNQPHTHTHLDLPTYPFQHHHYWLEL
                     PSAQTSPGQRRSRRSAPDTAESEFWDAVNEEDLQSLAETLDIDASALDTVVPALSAWH
                     RHQHDQARINTWTYQETWKPLTLPTTHQPHQTWLIAIPETQTHHPHITNILTNLHHHG
                     ITPIPLTVNHTHTNPQHLHHTLHHTRQQAQNHTTGPITGLLSLLALDETPHPHHPHTP
                     TGTLLNLTLPQTHTQTHPPTPLWYATTNATTTHPNDPLTHPTQAQTWGLARTTLLEHP
                     THTAGIIDLPTTPTPHTLHHLTQTLTQPHHQTQLAIRTTGTHTRRLTPTTLTPTHQPP
                     TPTPHGTTLITGGTGALATHLTHHLTTHQPTQHLLLTSRTGPHTPHAQHLTTQLQQKG
                     IHLTITTCDTSNPDQLQQLLNTIPPQHPLTTVIHTAGVNLFAPVSETDAESFSSVTAA
                     KATGAAILHELLLDHETLEHFILFSSGAGAWGSGNQCAYSAANAYLDALATHRQTHGL
                     PGASIAWGPWAGKGMSAGDAAHGYLEKRGILPMEPRMALAAFHRARAQRPNSNLIIAD
                     IDWERFVPAFTARRHSPLIEDIPEVRQAAQELEAAASTAKTTTAQPIATSLRERLARL
                     TSSKQNQVLLGLIRTGICTVLGLRNPEGIEDQRAFRDLGFDSLTSAQFSKELAKETGL
                     PLPPSLVFDYPTPQECAAHLRTQLVDLDDEEDAALSNALPQVAHRRTVEDEPIAIIGM
                     ACRFPGGVRSADDLWELLASGKDAIGVFPTDRGWDLDTLYDPDPDHPGTCYTRNGGFL
                     YGAGHFDAEFFGISPREALAMDPQQRLLLETAWETIEHAGINPHTLHGTPTGVFAGIN
                     AQDHAAHIRQSRDVETIEGYALTGSSGSVASGRVAYTLGLEGPAVSVDTACSSSLVAL

  Query Match             80.0%;  Score 1032.2;  DB 1;  Length 64957;
  Best Local Similarity   87.9%;  Pred. No. 4.4e-119;
  Matches 1137;  Conservative    0;  Mismatches  153;  Indels    3;  Gaps    1;

Qy          1 GAGCCGATCGCCATCGTTGGCATGGCCTGTCGTTTCCCCGGCGGAGTGACCTCGGCGGAC 60
              ||||| |||| ||| |||||||||| || || | ||||| || || |||| || ||
Db      44445 GAGCCCGTCGCGATCATTGGCATGGCGTGCCGCTATCCCGGAGGCGTCACCTCAGCCGAG 44386

Qy         61 GACTTCTGGGATCTGATCTCCTCCGAGCAGGACGCGATCGGCGGATTCCCCACCGACCGC 120
              ||  | ||||| ||| || | || | |  |||| || ||||||  || || || |||||
Db      44385 GAGCTGTGGGAACTGCTCGCATCGGGGAGGGACACGGTCGGCGAGTTTCCGACGGACCGT 44326

Qy        121 GGCTGGGACCTGGACACGCTCTACGACCCCGACCCCGACCACCCCGGCACCTGCTACACC 180
              || |||||||||||  | || | ||| || || || |  |  || |||||||  ||||||
Db      44325 GGGTGGGACCTGGAAGCACTGTTCGATCCGGAACCGGGTCGGCCGGGCACCTCGTACACC 44266

Qy        181 CGAAACGGCGGATTCCTCTACGACGCAGGCCACTTCGACGCCGAATTCTTCGGCATCAGC 240
              ||    ||  | ||||||||||||||| ||  | ||||||||||  ||||||||  |||||
Db      44265 CGCTGTGGGAGTTTCCTCTACGACGCGGGGGAGTTCGACGCCGGCTTCTTCGGGATCAGT 44206

Qy        241 CCCCGCGAAGCCCTCGCCATGGACCCCCAGCAACGACTCCTCCTCGAAACCGCCTGGGAA 300
              || || || || || || |||||||| |||||  ||| | || || ||  || | |||||
Db      44205 CCGCGTGAGGCACTGGCGATGGACCCGCAGCAGCGATTGCTGCTGGAGGCCTCATGGGAG 44146

Qy        301 ACCATCGAACACGCCGGCATCAACCCCCACACCCTCCACGGCACCCCCACCGGAGTCTTC 360
              |||| || || || || ||  ||||    ||| | | ||| | ||  || || || |||
Db      44145 GCCATGGAGCAGGCAGGTATTGACCCTACGACCGTACGCGGGAGCCAGACAGGCGTGTTC 44086

Qy        361 ACCGGCACCAACGGACAGGACTACGCACTTCGCGTGCAC---AACGCGGGCCAGTCAACC 417
              | |||  ||     |||| ||| | ||    | |||||  |||||  |       |||
Db      44085 GCGGGCCTCATTCCGCAGGCCTATGGACCCAGGCTGCACGAAAACGCCGCAGCCGACACC 44026

Qy        418 GATGGTTTCGCACTGACCGGAACCGCCGGCAGCGTCATCTCCGGTCGTATCTCGTACACG 477
              || || |  |  |||||||| ||  |||| || ||   ||||||||||||||||||||
Db      44025 GAGGGCTATGTCCTGACCGGCACATCCGGGAGTGTGGCCTCCGGTCGTATCTCGTACACG 43966

Qy        478 TTTGGTTTTGAGGGTCCTGCGGTGTCGGTGGACACGGCTTGTTCCTCGTCGTTGGTGGCT 537
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      43965 TTTGGTTTTGAGGGTCCTGCGGTGTCGGTGGACACGGCTTGTTCCTCGTCGTTGGTGGCT 43906

Qy        538 TTGCATCTGGCCTGTCAGGCGTTGCGTGCGGGTGAGTGCTCGATGGCGCTTGCCGGGGGT 597
              || ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      43905 TTACATCTGGCCTGTCAGGCGTTGCGTGCGGGTGAGTGCTCGATGGCGCTTGCCGGGGGT 43846

Qy        598 GTGACGGTGATGTCGTCTCCGGGTGCCTTCGTGGAGTTTTCGCGGCAGCGGGGTCTGGCC 657
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      43845 GTGACGGTGATGTCGTCTCCGGGTGCCTTCGTGGAGTTTTCGCGGCAGCGGGGTCTGGCC 43786

Qy        658 GCGGACGGGCATTGCAAGGCGTTCTCGGCGGCGGCGGACGGGACCGGCTGGGGTGAGGGT 717
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      43785 GCGGACGGGCATTGCAAGGCGTTCTCGGCGGCGGCGGACGGGACCGGCTGGGGTGAGGGT 43726

Qy        718 GTGGGGATGCTGCTGGTGGAGCGGCTCTCCGACGCCCATCGCAACGGTCACCGTGTCCTG 777
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      43725 GTGGGGATGCTGCTGGTGGAGCGGCTCTCCGACGCCCGTCGCAACGGTCACCGTGTCCTG 43666

Qy        778 GCCGTGGTGCGTGGCAGTGCGGTCAACCAGGACGGTGCGAGCAACGGTCTGACCGCGCCC 837
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      43665 GCCGTGGTGCGTGGCAGTGCGGTCAACCAGGACGGTGCGAGCAACGGGCTGACCGCGCCC 43606

Qy        838 AACGGGCCGTCCCAGCAGCGTGTCATCCGCCAGGCCCTCGCCAACGCCGGCTTGTCGGCC 897
              |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      43605 AACGGGCCCTCCCAGCAGCGTGTCATCCGCCAGGCCCTCGCCAACGCCGGCTTGTCGGCC 43546

Qy        898 GGTGATGTCGACGCGGTGGAGGCCCACGGCACCGGCACCACTTTGGGCGACCCGATCGAG 957
              |||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db      43545 GGTGATGTCGATGCGGTGGAGGCCCACGGCACCGGCACCACTTTGGGCGACCCGATCGAG 43486

Qy        958 GCCCAGGCCCTCCTCGCGACCTACGGACAGGACCGTGCCGGCGAGGGGCCGCTGTGGCTG 1017
              |||||||||||| |||||||||| ||||||||||||||||||||||||||||||||||||
Db      43485 GCCCAGGCCCTCCTTGCGACCTACGGGCAGGACCGTGCCGGCGAGGGGCCGCTGTGGCTG 43426

Qy       1018 GGCTCGGTCAAGTCCAATGTCGGTCACACACAGGCTGCCGCGGGCGTCGCCGGGGTGATC 1077
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      43425 GGCTCGGTCAAGTCCAATGTCGGTCACACACAGGCTGCCGCGGGCGTCGCCGGGGTGATC 43366

Qy       1078 AAGATGGTGATGGCGCTGCGGCATGGTCTGCTGCCGCGGACGTTGCATGTGGATGAGCCG 1137
              |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db      43365 AAGATGGTGATGGCGCTGCGGAATGGTCTGCTGCCGCGGACGTTGCATGTGGATGAGCCG 43306

Qy       1138 TCGCCGCATGTGGACTGGTCCGCGGGTGCGGTGCAGCTGCTGACGGAGACGGTGCCCTGG 1197
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      43305 TCGCCGCATGTGGACTGGTCCGCGGGTGCGGTGCAGCTGCTGACGGAGACGGTGCCCTGG 43246

Qy       1198 CCCGGCGGGGAGGGGCGGCTACGGCGGGCAGGAGTGTCATCATTCGGCGTCAGCGGCACC 1257
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      43245 CCCGGCGGGGAGGGGCGGCTACGGCGGGCAGGAGTGTCATCATTCGGCGTCAGCGGCACC 43186

Qy       1258 AACGCCCACGTCATCCTCGAAGAAGCACCCGCC 1290
              ||||||||||||||||||||||||||||||||
Db      43185 AACGCCCACGTCATCCTCGAAGAAGCACCCGCC 43153
```

```
RESULT 7
AB070949
LOCUS       AB070949               78210 bp    DNA     linear   BCT 11-OCT-2001
DEFINITION  Streptomyces avermitilis polyene macrolide biosynthetic gene
            cluster.
ACCESSION   AB070949
VERSION     AB070949.1  GI:15824136
KEYWORDS    .
SOURCE      Streptomyces avermitilis DNA.
  ORGANISM  Streptomyces avermitilis
            Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
            Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1
  AUTHORS   Omura,S., Ikeda,H., Ishikawa,J., Hanamoto,A., Takahashi,C.,
            Shinose,M., Takahashi,Y., Horikawa,H., Nakazawa,H., Osonoe,T.,
            Kikuchi,H., Shiba,T., Sakaki,Y. and Hattori,M.
  TITLE     Genome sequence of an industrial microorganism Streptomyces
            avermitilis: deducing the ability of producing secondary
            metabolites
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 98 (21), 12215-12220 (2001)
  MEDLINE   21477403
```

```
REFERENCE   2  (bases 1 to 78210)
  AUTHORS   Ikeda,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-SEP-2001) Haruo Ikeda, Kitasato University, Dept. of
            Microbial Chemistry, School of Pharmaceutical Sciences; 5-9-1
            Shirokane, Minato-ku, Tokyo 108-8641, Japan
            (E-mail:ikeda@mc.pharm.kitasato-u.ac.jp, Tel:+81-3-5791-6242,
            Fax:+81-3-3444-6197)
FEATURES             Location/Qualifiers
     source          1. .78210
                     /organism="Streptomyces avermitilis"
                     /db_xref="taxon:33903"
                     /note="Actinobacteridae"
     CDS             complement(1. .1254)
                     /codon_start=1
                     /transl_table=11
                     /product="putative lipoprotein"
                     /protein_id="BAB69301.1"
                     /db_xref="GI:15824137"
                     /translation="MGEALITLCTRSTRLAATAGVLGGVLAVTALGGTSNAATSGNGA
                     GNHPTSQARINQAGAQEASDARIEITPGQGTYSVGTNHPVSVTVTSGKLTKVTMTAVA
                     TGAEIPGTLSTDGTSWKPNGRLERATKYQIAAEAGDVKGRSTTGSATITTVSPANDFI
                     GHLSPEDGSTVGVAMPVTVNFDKAISDKAAVQSEIRVSSSSGQRVVGHWFNDHRVDFR
                     PENYWRPGSTVTVTLNRHGIQKTVTFKIGRSQISTVDAKTKQMTVVRDGKTIKTIPIS
                     SGSPEHPTYNGQMVISEKFRQIHMNGASVGLTEKNGKPSYDIKAVPHAMRLTDSGTFI
                     HGNYWGTDSVFGEANTSHGCVGLKDVRGDSHGRQPAAWFFDHSIVGDVVIVKNSEEKT
                     MLAPDNGLSDWNMSWSEWVAGSPAT"
     CDS             complement(1256. .2290)
                     /codon_start=1
                     /transl_table=11
                     /product="putative transposase"
                     /protein_id="BAB69302.1"
                     /db_xref="GI:15824138"
                     /translation="MPVEFLSHDQVSRYGRFASAPTPGELEQFFRLDAKALESARSKR
                     APANRLGWAVQWGCVRMLGVFPTEDLSVVPEVVVRFAAEQLGVDPGFEAAYGARRQNR
                     YEHAWEIRDAYAYRESPAAEAQVREFLATRVWASWEGPRALFDRAVGWLVDNRVLLPG
                     ITTLTRLVAEVRAQENAALYRTLDEAVPEKLRQSMRDLLKVPEGRRVSELERLRRSPM
                     GVSGSADRLAFSERIRMEAVAMFTAGPGSTDIAKELRVSVQSVQRWRSGLAGDRPGRG
                     AVPQPVSVELSETLFAALDEERAKGFAPSWTIRYLHHDFPIFGQRYAYKWSQGCLHSF
                     GRSAQGYRDQ"
     gene            2935. .21081
                     /gene="pteA1"
     CDS             2935. .21081
                     /gene="pteA1"
                     /codon_start=1
                     /transl_table=11
                     /product="modular polyketide synthase"
                     /protein_id="BAB69303.1"
                     /db_xref="GI:15824139"
                     /translation="MVESELNATGNRDCSASADSADGIAVIGLACRLPGAADPAAFWR
                     LLSEGADAITDVPPDRWDGAVVADASAPDRTGIRRGGFLDSIGDFDAGFFGLSPKEAA
                     AMDPQQRLVLELAWEALEDAHVRAATLRSTRTGVFVGAMWDDYATLHHRSGLTAISPH
                     TVTGLHRSIIANRVSYFLGLNGPSLTVDSGQSSSLVSVHLACESLRKGESTIALAGGV
                     NLNIVPESALGAAKFGGLSPDGRCFTFDARANGYVRGEGGGIVVLKPLARAVADGDPV
                     YCVIRGSAVNNDGGGDGLTVPLQSGQEQVLRLAYERAGVNPADVGYVELHGTGTKVGD
                     PIEAAALGAVLGTGREPGRPLRVGSAKTNVGHLEGAAGITGLLKAVLSLNHRELPASL
                     NFGTPNPAIPLDRLNLRVQTEHSEWAAEDGRPLLAGVSSFGMGGTNCHIVLASGAAVE
                     AGEGATAGGGAAPTFWPLSAKTASGLRAQAAALLAHTEAPLGLALPDVGWSLATGRTA
                     FEHRAVVVGEVRDDFLRALRELSTGGIDAALTTGRAGPRGELAFLFSGQGSQRAGMGR
                     ELAESVPAFASALDEVCAHLDPLLPRPLREVMFAPVGTGEAAELDQTLYTQTSLFAVE
                     VALFRLLESWDITPDVLMGHSIGELAAAHVAGVLSLADACALVAARGRLMQALPAGGA
                     MIAVQAAEDEVRARIGERTDRVSIAALNGPDSVVVSGDEDLATEIADAFAALGRKTSR
                     LRVSHAFHSPHMEPMLAEFRNVAEGLTFHAPRIPIVSNVTGRLSEEPSGYEGCTAAYW
                     VRHVREAVRFADGVARLDEQGVRTYLELGPGGVLASMARAGADSESLFVPALRTRRPE
                     AQALLAAVASLHVHGVEPDWDALFGGGRGARRKVALPTYAFERQRYWLDGDTTSPVAT
                     TGWAPVAPQPDAIAETDSAPVGALGRRLAGLAESERDEAVLDMVRAHIAAVLGHAEAR
                     QVETEWTFKDLGFDSLSSVELRNQLSEATGLRLPSGLLFDHPTPAALARHLAAQSLGG
                     GDAAAAAPATVADPHEPIAIVAMSCRLPGGVSSPEDLWRLLASGGDGISGFPTDRGWD
                     VETLYDPEPGTPGKSSTRHGGFLHEAAEFDAGFFGISPREAAAIDPQQRLVLETAWEA
                     FERAGIDPATLRGSRAGVFIGATAQDYGPRLHEPADGVEGYLLTGTTTSVASGRVAYS
                     FGLEGPAVTVDTACSSSLVALHLAVQSLRSGECTMALAGGVTVMSTPGMFVEFSRQRG
                     LSADGRCKAFAAGADGTGWAEGVGMLLVERLSDARRNGHQVLAVVRGSAVNQDGASNG
                     LTAPNGPSQQRVVRQALAAAGLAPADVDAVEAHGTGTLLGDPIEAEALIAAYGQGRDG
                     DRPLWLGSLKSNIGHTQAAAGVAGVIKMVMAMRHGVLPRTLHVDEPTPHVDWSASGVR
                     LLTEAVEWPARDLPRRAAVSSFGISGTNAHVIVEQAATSEEPARGTAVAGPVPWVVSA
                     KSDVALREQAGRLLSFLDAEGVGGAGVRPEDVALSLATSRAVLERRAAVVGGDLAELR
                     QGLEALASGASGAAPVGGAVGGRVGFLFSGQGSQRIGMGRELYAAYPVFAAAYDEVCA
                     FLDAPVDVDSEELNQTGTTQPALFAVEVALFRLLESWGIRPDYVAGHSVGEIAAAHVA
                     GVLSLEDAAKLVSARAALMQALPDGGAMVAVQATEDEVLRHLTDGVGIAAVNGPRSVV
                     VSGAEDAVTAVGEIFAGRGRKTSRLKVSHAFHSPLMDPMLEEFAAVVRGLTFSESRIP
                     VVSNLTGRLAEVYTPEYWVRHVREAVRFADGIETLGELGVTTFVEVGPGGVLSALAQG
                     CFDDGTDIVTVPVLRADRPEPYALVTAIGRLHTHGISPDWQVLFPGARRVDLPTYAFQ
                     RERYWLDAPRAATAGAPDAMDAEFWDSVESEDRASLGALLGLEPTELDVVASKLSAWR
                     RKRREQSTTDGWRYRITWQPLDDMVRAAPSGTWLYVVSEETAWTEAIRAGLTELGVAL
                     VPFAVGEDTDRDALARAFADAGQEQADGILFAAAQGEAGANALQRLVLLVQALGDAGI
                     EAPLWCLTSGAVSTGTADLLTDPVSAQLWGLGRIVALEQPQRWGGLVDLPAELDSRAL
                     ERLAGLLAQADEDQLAVRASGVYGRRLVRAPLSAAPADAPWSPRGTVLVTGGTGALGG
                     QVARWLAGAGAEHLVLTSRRGPDAPGAAELKAELEELGAQVTVVACDAADRDALAELF
                     GRYIVNAVVHTAGVLDDGLVESLTPERLNHVLRPKVDAALHLHELTRDRQDLDAFVLF
                     SSMTGVWGNGGQGAYGAANAFLDALAEHRRAQGVPALAVAWGSWADGGMADGAAGDHL
                     RRRGVRAIAALPAISVLHGALTHGETSMTVADVDWERFVPAFVGTRPCPLLLGVPEAR
                     EALEARARAAQSPEVPVSALVQRLLGATPGEQGRILLDLVREQAATVLGHTGKGAFEA
                     DRTFRETGFDSLTAVELRHRLNTATGLRLPTTLVFDHPTPTALARHLGEELLGRHGLE
                     TAAAQATAAAAADEPLAIVGMSCRLPGGVQSPEDLWELLSSGRDAMSGFPTDRGWDID
                     ALYDPDPDRPGKTYSREGGFIQGADRFDAGLFGISPREALAMDPQQRLLLETAWEAFE
                     RAGIALASVRASRTGVFIGTNGQDYANGLRNAPEEIEGYALTGKAASVVSGRISYTFG
                     LEGPAVTVDTACSSSLVALHLAAQALRSGECSMALVGGVTVMTTPDLFVEFSRQRGLS
                     PDGRCKAFAAGADGTGWGEGVGLLLVERLSDARRNGHQVLALVRGSAVNQDGASNGLT
                     APNGPSQQRVIRQALASAGLTPADVDAVEAHGTGTRLGDPIEAQALLATYGQEHSDDR
                     PVWLGSVKSNIGHTQAAAGVAGVIKMVLAMRHGVLPQTLHVDEPTAHVDWSAGAVRLL
                     TEPVQWPGTTGPRRAAVSSFGIGGTNAHTIIEEAPAAAAAAEPAREHRPVPVPWVLSAK
                     SEVALRAQAERLLAFATDDVSPVDAGFSSATTRSALEHRAAVIGSDPAELRAGLEALA
                     AGEPAANVVAGRAHSADKVGFLFSGQGSQRLGMGRELYAAYPVFAAAYDEVCALLDAT
                     VDVDAETLHQTGSTQPALFAVEVALFRLLESWGIRPDYVAGHSVGEIAAAHVAGVLSL
                     EDAAKLVSARAALMKALPAGGVMVAVQAAEDEMLPHLIDRAGIAAINGPRSVVISGAE
                     DAVTAIAEVFAAQGRKTSRLKVSHAFHSPLMDPMLEEFAEVVRGLTFSEPWIPVVSNL
                     TGRLTESYTPEYWVRHVREAVRFADGVLTLHELGVTTFVEIGPGGVLSALVQGCLDDG
                     TDIVTIPALRADRPEPQALTTAYAQLHVSGAEVDWHAFFPGARRVELPTYAFQRERYW
                     LDAPASAGDMRAVGQGEAGHPLLGAAVPLADSVGHLLTGRLSKHTHPWLVDHAVSGVA
                     LLPGTAFVELAITAADAVGCDLLEELTLETPLLVPERGGVALQVHVGADDGSGRRPLT
                     VHSQIDDGDAYAHEDDSERAWVRHASGVLTVATGGPVDGSLTAWPPVGAEPVDVDGFY
                     DRLAGMALDYGPVFQGLRAAWRSGDDFFAEVELPGAERTDAGAYGLHPALFDAALHTV
                     WLGAVEPEAGTGNGLLPFAWSGVRLAAAGASDLRVKVSRAGTGTSTVSLVLADGTGGP
                     VAEVDSLTLRPVPADQLRAGSGSDDAVFGLKWTPVSLPSAPDGMRIETYEDLAALRAT
                     DPPADTEADAAAAADAVVVPCPTGTGADLATRVREVTHAVLELVQWWLAEERPGRLVL
                     VTRGGDLAQSAVCGLVRSAQTENPDRIVLVETSSDAHAGAGSGVDDADLARVLPGLIA
                     SGEPQAAVHADDEVRVPRLVRVAVPDAEPELPGLGSGTVLLTGASGGLGGLFARHLVA
                     EHGVRSLLLVSRRGGDAPGAAELTADLTAQGADVTWAACDVADRAAVRALLAGPGQSL
                     SAIIHTAGVLDDGIIGSLTPERLDAVFRPKVDAALNLHDLAGELSGDLSAFVLFSSVA
                     GTLGTPGQANYAAANTFLDALAEHRRAAGQRATSLAWGLWAQDSESAMTGSLDHTDLT
                     RIQRMGLAAIPSSDGLRLFDAALALDRAVVAPLRLDLGGLQGTVPPVLRGLRGTVRGT
                     ARRTAQTAPKGSLGQRLAGLAEAEREQVVLDLVRTEVAAVLGHAGAQSVGADDSFKDI
                     GFDSLTAVELRNRLNSAVGMRLPATLIFDYPNPLTLAHFLTGEAAGGEAATAVPTTLS
                     TATGTADEPIAIVGMACRYPGGVQSPEDLWQLVFSGRDAVSGFPEGRGWDVEQLYDPN
                     PDQWGTSYTREGGFLHDAADFDAEFFGISPREALAMDPQQRLLLETSWEAFERAGIDP
                     ATVRGSRTGVFAGVMYHDYGGRVHTSPPGLEGYLVNGSAGSIASGRVSYTFGLEGPAV
                     TVDTACSSSLVALHLAAQALRSGECSMALVGGVTVMAGPSVFVEFSRQRGLSADGRCK
                     AFGAGADGTGWAEGVGMLLVERLSDARRNGHKVLAVVRGTAVNQDGASNGLTAPNGPS
                     QQRVIRQALANAGLSSDEVDAVEAHGTGTRLGDPIEAQALIATYGQERSEDRPLWLGS
                     LKSNIGHSQAAAGVGGVIKMVMAMQHGVLPQTLHVDEPTPHVDWSAGAVQLLTEAVEW
                     PASDGPRRAAVSSFGVSGTNAHVIVEQAATSEVTARGTAVAGPVPWVLSAKSDAALRE
                     QAGRLLSFLDAEGVGGAGVRPEDVAYSLATSRAVLECRAAVVGEELGEFRQGLEGIVS
                     GAAPVGGAVGGRVGFLFSGQGSQRIGMGRELYAAYPVFAAAYDEVCALLDATVDVDAE
                     TLHQTGSTQPALFAVEVALFRLLESWGIRPDYVAGHSVGEIAAAHVAGVLSLEDAAKL
                     VSARAALMQALPAGGAMVAIQATEAEVLPHLTHGVGLAAINGPRSVVLSGSEDAVLDI
                     AEVFAQQGRKTSRLKVSHAFHSPLMDPMLEEFAEVVRGLTFSEPRIPVVSNLTGRLAE
                     VYIPEYWVRHVREAVRFADGLHTLHELGVTMFVEIGPGGVLSALVQGCLDDGTHIVTV
                     PALRADRPEPYALVTAVGQLHTHGISPEWQALFPGARRVDLPTYAFQRERYWLDAPQA
                     SGDVRAAGLGAADHPLLGAAIATADSDGAVLTGLLSLGTHPWLADHAVNGIVLLPGTA
                     FVELAIRAGDETGCGWIEDLTLELPLVVPERGGVTLQVAVGPEDESGRRQLSVHSRVD
                     DRSWLRHATGVLSSAEAPSAAKLGAWPPAGAEAVDLTAFYEGMAAAGLDYGPVFQGLR
                     SVWRDGDDVFAEVELPDESAAEAAAFALHPALLDAALHALAAGGLVSLDDSPALPFAW
                     SGVVVQAAGAAMVRVKLSRTATGSVTLAVADGAGDPVASVESLSLRAASAEQLPGAVG
                     GDQLFGLEWTRFTLPSATDALMIETVADFDALRESEGTPDVVVVPVVPCPAGSGRETA
                     DRVHSTTSEVLALVQWWLAEERAGRLALVTRPGDLAHAAVWGLVRSAQSENPDRIVLV
                     EAEETDEAVRVLPAALASGEPQFAVRDGEVFVPRLVKATRSATGTATGTPDFGVGPVL
```

```
                     LTGASGALGGLVARHLVAEHGVRSLLLLSRRGAEAPGAVELEAELAAWGAEVRWAACD
                     VADREAVSGMLNGLGEQSLSAVVHTAGVLGDGIVASLTPERMREVFRPKVDAVLNLHE
                     CTRDMGLAAFVVFSSVAGMVGSAGQASYAAANSFLDAFSAHRRREGLPAISLAWGVWE
                     QSGAMTDGLVEADRARMARSGVLPLPPEEGLRLFDAALASDEAVLAPVRIDTGALRAG
                     EAPPVLRALVPAAAARRTARAAAAPSASSLAERLEEMPEAEREKTVLDLVRAEVAAGL
                     GHASDRTVRPEHAFQDLGFDSLTAVELRNRLNKATGLRLPATLVFDHPTPALLARHVF
                     MAAAGAAEPALVDSLLADLDRVEQELVTTLAASEARDRILSKLQGLLAIAGEFGKAPA
                     QEGPGTGGSDLESATDDEVFDLLGKEFGIS"
     gene            21273. .31847
                     /gene="pteA2"
     CDS             21273. .31847
                     /gene="pteA2"
                     /codon_start=1
                     /transl_table=11
                     /product="modular polyketide synthase"
                     /protein_id="BAB69304.1"
                     /db_xref="GI:15824140"
                     /translation="MSCRYPGGVDSPEDLWELVAGGRDAVSEFPADRGWDLDALYDPD
                     PESRGTSYARERGAFLRDLARFDAGLFGISPREALAMDPQQRLLLETSWELFERAGIDA
                     ASLAGSRTGVFAGVMNHEFLAALQNSPQDLEGYLGTGTSGSVASGRVAYTFGLEGPAV
                     TVDTACSSSLVALHLAVQSLRNGECSLAVAGGVTAMVGPATFVGFSRQRGLAPDGRCK
                     PFAAGADGTGWGEGAGLLLVERLSDARRNGHEVLAVVRGSAVNQDGASNGLTAPNGPS
                     QQRVIRQALAAAQLTTAQVDAVEAHGTGTRLGDPIEAQALLATYGQGRPDDKPLWLGS
                     IKSNIGHTQAAAGVAGIIKMVMAMRHGELPRTLHVDAPSVHVDWTAGAVRLLTERTPW
                     SASDQPRRAGISSFGISGTNAHTIIEEAPSAEAFAQDTGAPEGPEEGAESAEPVASEA
                     ARPAAGGPVPWVISGRTEEALRAQAERLLSRTDAEVPPLDVAFSLATTRTALEHRAVL
                     QGETPAELAEGLRAVASGSPARGVARGLSGTGGRVGFLFSGQGSQRLGMGRELYASYP
                     AFAAAYDEVCAQLDAVVDVDSEELHRTGAAQPALFAVEVALFRLLESWGIRADYVAGH
                     SVGELAAAHVAGVLSLDDAAKLVSARARLMQALPAGGAMVAVQATEDEVLPHLTDAVG
                     IAAINGPRSVVVSGAEDAVTAIAEAFRQQGRKSSRLKVSHAFHSPLMDPILEEFAEVV
                     RSLTFHRPQIPVVSNLTGRLAEPYTPQYWVRHVREAVRFADGVRTLNDLGVTTFVEIG
                     PGGVLSALAEGCLDDDVVTVPALRADRPERVALVAAVSELHAHGVSPDWHAFFPGARR
                     VDLPTYAFQYERYWLDASPAAPGGVRAAGLGSADHPLLGAAVSLAGGDERLLTGRLSL
                     RTHPWLSDHAVLGTALLPGTAFVELAVRAADEAGCALLEDLTLEAPLVVPEQGGVAVQ
                     VWVGAVDGTGRRPLTVHSRPEDDSDLPWVRHATGFVTEGSEEAGAAAADRTLTAWPPT

  Query Match          60.8%;  Score 784.4;  DB 1;  Length 78210;
  Best Local Similarity  76.0%;  Pred. No. 1.6e-88;
  Matches  981;  Conservative    0;  Mismatches  306;  Indels    3;  Gaps    1;

Qy          1 GAGCCGATCGCCATCGTTGGCATGGCCTGTCGTTTCCCCGGCGGAGTGACCTCGGCGGAC 60
              |||||  |||| |||||  ||||||  |||  ||  ||||||||||  ||     |||  ||||
Db      10588 GAGCCCCTCGCGATCGTCGGCATGAGCTGCCGCCTCCCCGGCGGCGTCCAGTCGCCGGAG 10647

Qy         61 GACTTCTGGGATCTGATCTCCTCCGAGCAGGACGCGATCGGCGGATTCCCCACCGACCGC 120
              |||  |  |||||  |||  |    |||||   |  |||||||||    |||  ||||||||||||||
Db      10648 GACCTGTGGGAGCTGCTGAGCTCCGGCCGGGACGCGATGTCCGGCTTCCCCACCGACCGC 10707

Qy        121 GGCTGGGACCTGGACACGCTCTACGACCCCGACCCCGACCACCCCGGCACCTGCTACACC 180
              |||||||||  |  |||  ||||||||||||||||||||||||||  |||||||     |||  |
Db      10708 GGCTGGGACATCGACGCGCTCTACGACCCCGACCCCGACCGGCCCGGCAAGACGTACTCG 10767

Qy        181 CGAAACGGCGGATTCCTCTACGACGCAGGCCACTTCGACGCCGAATTCTTCGGCATCAGC 240
              ||   |  |||||  |||  ||  |  |||  |  ||  ||||||||||    |||||||  |||
Db      10768 CGCGAGGGCGGCTTCATCCAGGGCGCCGACCGCTTCGACGCCGGTCTCTTCGGGATCTCG 10827

Qy        241 CCCCGCGAAGCCCTCGCCATGGACCCCCAGCAACGACTCCTCCTCGAAACCGCCTGGGAA 300
              ||  ||  ||  ||  ||  ||  ||||||||  ||||||||  ||  ||  ||  ||  |||||  |||||
Db      10828 CCGCGTGAGGCGCTGGCGATGGACCCGCAGCAACGGCTGCTGCTGGAGACCGCGTGGGAG 10887

Qy        301 ACCATCGAACACGCCGGCATCAACCCCCACACCCTCCACGGCACCCCCACCGGAGTCTTC 360
              ||  ||||  |  ||||||||||   ||  |     ||  |||  ||  ||  ||  ||||||  ||||||
Db      10888 GCCTTCGAGCGCGCCGGCATCGCCCTCGCGTCCGTCCGCGCCAGCCGCACCGGTGTCTTC 10947

Qy        361 ACCGGCACCAACGGACAGGACTACGCACTTCGCGTGCACAACGCGGGCCAGTCAACCGAT 420
              |  ||||||||||||  ||||||||||      |  |  |  |||||||  |  ||    |  |||
Db      10948 ATCGGCACCAACGGCCAGGACTACGCCAACGGGCTACGCAACGCGCCCGAGGAGATCGAG 11007

Qy        421 GGTTTCGCACTGACCGGAACCGCCGGCAGCGTCATCTCCGGTCGTATCTCGTACACGTTT 480
              ||  |  |||  ||  |||||  |   ||  |  |   |||  |||||||  ||  ||||||  ||||||||
Db      11008 GGCTACGCGCTCACCGGCAAGGCGGCCTCGGTCGTCTCCGGACGCATCTCCTACACGTTC 11067

Qy        481 GGTTTTGAGGGTCCTGCGGTGTCGGTGGACACGGCTTGTTCCTCGTCGTTGGTGGCTTTG 540
              ||  |  ||||||||  |||||  ||||  ||||||||  ||  ||  ||  |||  |  |||||  ||
Db      11068 GGCCTCGAGGGTCCCGCGGTCACGGTCGACACGGCCTGCTCGTCTTCGCTCGTGGCCCTG 11127

Qy        541 CATCTGGCCTGTCAGGCGTTGCGTGCGGGTGAGTGCTCGATGGCGCTTGCCGGGGGTGTG 600
              ||  ||||||   |||||  ||||   ||  ||||||||||||||  ||  |  ||||||  ||
Db      11128 CACCTGGCCGCCCAGGCACTGCGGAGCGGCGAGTGCTCGATGGCCCTGGTCGGGGGCGTC 11187

Qy        601 ACGGTGATGTCGTCTCCGGGTGCCTTCGTGGAGTTTTCGCGGCAGCGGGGTCTGGCCGCG 660
              ||  ||  |||  |  |  ||||      ||||||||||     ||||||||  ||  |||  |  ||
Db      11188 ACCGTCATGACCACGCCGGACCTGTTCGTGGAGTTCAGCCGGCAGCGCGGGCTGTCGCCG 11247

Qy        661 GACGGGCATTGCAAGGCGTTCTCGGCGGCGGCGGACGGGACCGGCTGGGGTGAGGGTGTG 720
              |||||  |  ||||||||||||  ||||  |  ||  |||||  ||  ||||||||  |||||  |||
Db      11248 GACGGCCGGTGCAAGGCGTTCGCGGCCGGTGCCGACGGCACGGGCTGGGGCGAGGGCGTG 11307

Qy        721 GGGATGCTGCTGGTGGAGCGGCTCTCCGACGCCCATCGCAACGGTCACCGTGTCCTGGCC 780
              ||  ||  ||||||||||  |||  |  |||||||||  |||  |||||  ||||   ||  |||||
Db      11308 GGCCTGTTGCTGGTGGAACGGTTGTCCGACGCCCGTCGGAACGGGCACCAGGTGCTGGCG 11367

Qy        781 GTGGTGCGTGGCAGTGCGGTCAACCAGGACGGTGCGAGCAACGGTCTGACCGCGCCCAAC 840
              ||||  ||  ||||||||||||  ||||||||||||||||  ||||||||||||||||||||
Db      11368 TTGGTCCGCGGCAGTGCGGTGAACCAGGACGGTGCGTCCAACGGTCTGACCGCGCCCAAC 11427

Qy        841 GGGCCGTCCCAGCAGCGTGTCATCCGCCAGGCCCTCGCCAACGCCGGCTTGTCGGCCGGT 900
              ||  ||  ||  ||||||||  ||  |||||  ||||  ||  ||    ||||||  |  ||  |  |
Db      11428 GGCCCCTCGCAGCAGCGGGTGATCCGGCAGGCGCTGGCGTCCGCCGGGCTCACGCCGGCC 11487

Qy        901 GATGTCGACGCGGTGGAGGCCCACGGCACCGGCACCACTTTGGGCGACCCGATCGAGGCC 960
              ||  ||  |||||  ||  ||||||||||||||  |||||     |  ||||||||||||||||
Db      11488 GACGTGGACGCCGTCGAGGCCCACGGCACGGGCACGCGGCTCGGCGACCCGATCGAGGCG 11547

Qy        961 CAGGCCCTCCTCGCGACCTACGGACAGGACCGTGCCGGCGAGGGGCCGCTGTGGCTGGGC 1020
              |||||  ||||||||  ||  |||||  |||||  |    |||  |||  |||||  ||||||||||
Db      11548 CAGGCGCTCCTCGCCACGTACGGCCAGGAACACTCCGACGACCGGCCGGTGTGGCTGGGC 11607

Qy       1021 TCGGTCAAGTCCAATGTCGGTCACACACAGGCTGCCGCGGGCGTCGCCGGGGTGATCAAG 1080
              ||||||||||||  ||||||||||  |||||  ||||||||  ||||||||  |||||||||
Db      11608 TCGGTCAAGTCCAACATCGGTCACACGCAGGCCGCCGCGGGTGTCGCCGGTGTGATCAAG 11667

Qy       1081 ATGGTGATGGCGCTGCGGCATGGTCTGCTGCCGCGGACGTTGCATGTGGATGAGCCGTCG 1140
              ||||||  |||||  ||||||  ||  ||||||||  ||||  |  ||||||||  ||||||  |
Db      11668 ATGGTGCTGGCGATGCGGCACGGCGTGCTGCCGCAGACGCTCCATGTGGACGAGCCGACC 11727

Qy       1141 CCGCATGTGGACTGGTCCGCGGGTGCGGTGCAGCTGCTGACGGAGACGGTGCCCTGGCCC 1200
              ||||  ||  ||||||||  ||  ||  ||  ||||  |||||||||  |||  ||||||  |||||
Db      11728 GCGCACGTCGACTGGTCGGCAGGCGCCGTGCGGCTGCTGACCGAGCCGGTGCAATGGCC- 11786

Qy       1201 GGCGGGGAGGGGCGGCTACGGCGGGCAGGAGTGTCATCATTCGGCGTCAGCGGCACCAAC 1260
                 ||||       |||  ||  |||||  |  |||||  ||  |||  ||  ||||||||||
Db      11787 --CGGGACCACCGGGCCCCGCCGGGCCGCGGTGTCCTCCTTCGGTATCGGCGGCACCAAC 11844

Qy       1261 GCCCACGTCATCCTCGAAGAAGCACCCGCC 1290
              ||  |||   |||  ||||  |||||  ||||||
Db      11845 GCGCACACGATCATCGAGGAAGCGCCCGCC 11874


RESULT 8
SNA278573/c
LOCUS       SNA278573               84985 bp    DNA     linear   BCT 17-MAY-2001
DEFINITION  Streptomyces natalensis pimaricin biosynthetic gene cluster.
ACCESSION   AJ278573
VERSION     AJ278573.1  GI:12055067
KEYWORDS    ABC transporter; cholesterol oxidase; cytochrome P450 monoxygenase;
            ferredoxin; glucosyl transferase; macrolide efflux pump; mycosamine
            dehyralase; orfX; pimA gene; pimB gene; pimC gene; pimD gene; pimE
            gene; pimF gene; pimG gene; pimH gene; pimI gene; pimJ gene; pimK
            gene; pimS0 gene; pimS1 gene; pimS2 gene; pimS3 gene; pimS4 gene;
            polyketide synthase; sensory transduction protein; thioesterase.
SOURCE      Streptomyces natalensis.
  ORGANISM  Streptomyces natalensis
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1  (bases 1 to 84985)
```

```
  AUTHORS   Aparicio,J.F., Fouces,R., Mendes,M.V., Olivera,N. and Martin,J.F.
  TITLE     A complex multienzyme system encoded by five polyketide synthase
            genes is involved in the biosynthesis of the 26-membered polyene
            macrolide pimaricin in Streptomyces natalensis
  JOURNAL   Chem. Biol. 7 (11), 895-905 (2000)
  MEDLINE   20547809
   PUBMED   11094342
REFERENCE   2  (bases 1 to 84985)
  AUTHORS   Aparicio,J.F.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-JUN-2000) Aparicio J.F., Molecular Genetics,
            Institute Of Biotechnology - Inbiotec, AV. REAL 1 - LEON, 24006,
            SPAIN
FEATURES             Location/Qualifiers
     source          1. .84985
                     /organism="Streptomyces natalensis"
                     /db_xref="taxon:68242"
     CDS             <1. .2086
                     /note="ORFX"
                     /codon_start=2
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="CAC20917.1"
                     /db_xref="GI:12055068"
                     /translation="TRPDAFREATEPYGALLAPLMPERSADRVEGTKWTSDWSQARFL
                     AHDAVCEVLLTLAGQHPLVLLLEDLQWADTASLDLLRLLSPRCQGHPLGIVLTARKFE
                     TESDATLRRIVSDVLRGPRTATLRLCGLPRQAVGALVEAHAGSGVDPEVVEVLHRRSE
                     GNPYFVMQLLSLVGDARRLHRPDAVDALLAHIPTGVREALNQRFAALPEPVLRVLRLC
                     AVIGAEVDTDLLRRTATCDASVNAGLESAIRAGLLGEDPHHPGQLHFAHALVQETLFD
                     ELPRADRRRLHAKVADALYARRRGHVGDEEIERVAHHSWHAKSALPAAEVLPRLLLAA
                     EHAEQAMAYEQVEIWLRRAVHLVGFLPPDDPATVRLEQKLHIQLGQVLATTRGYGHPE
                     AETALTRGRALSTVTHAPEDPSVLWALCAASLVTGRYDASRQLSGLLRDLSWQTREPV
                     AVLGAAYGEGIVLHVRGMLPQALAELEHGVERADHFANEGHALARTFQHDPRVSCRSY
                     NTFTHWLLGHRQIADARRRELLSLTEYESRPSDRSFALYVDGVVAAWEGDTRTALASS
                     DEGVRVAGDHGLLYWKAMLGVVKGWGLVHAGDRDEGLALMHTSLAELGQSRTYLRHPL
                     HLGLLGQAQHHNGQIEEARTTFRMLLAAVEQRRERVYLHPALPATPLLHELLGHGADE
                     ALADRDPGRGQV"
     gene            2190. .3566
                     /gene="pimK"
     CDS             2190. .3566
                     /gene="pimK"
                     /function="putative glucosyl transferase"
                     /codon_start=1
                     /transl_table=11
                     /product="PimK protein"
                     /protein_id="CAC20918.1"
                     /db_xref="GI:12055069"
                     /translation="MESARRPILFVSLPESGLLNPLLVLAGELSRQGVEDLWFATDEP
                     RRNDVKRIAEGSPVEFASLGEVDSEMSAVTWSDEVYREVTQPSRFKAHRAVVRHTYRP
                     GLQAEKFRRLQAVIDEVQPALMVIDCISGFAVDAAIARNIPYVLSVPFLPSNVLTAHT
                     HFAKSYTPRGFPVPHTGLSRRMTLAQRVANELFKLRTFAMFLNPRLGKVLAEDNRRRN
                     ELGLPKASFMARIEHADLVLCNSLAELDYPFDIPEKMRLVGAMVPPLPEAPDDQDLSR
                     WLDAQSSVVYVGLGTITRLTREQVGSMVEVARRLEDRHQVLWKLPSEQQHLLPPRESL
                     PGNLRVESWVPSQMDVLAHPHVKVFFTHGGGNGFNEGMYFGKPLVVRPLWVDCYDQAV
                     RGQDFGLSLTLDRPQTIDVNDVVDKLTRVLGTPSFYEKAERRAALMRSAGGRETAAGL
                     VLSLPALA"
     gene            complement(3662. .9736)
                     /gene="pimS4"
     CDS             complement(3662. .9736)
                     /gene="pimS4"
                     /function="polyketide synthase"
                     /codon_start=1
                     /transl_table=11
                     /product="PimS4 protein"
                     /protein_id="CAC20919.1"
                     /db_xref="GI:12055070"
                     /translation="MPNDKPNDKKLVDYLKWVTNDLHQTRQRLREVESAKQEPIAIVG
                     MSCRLPGGVRSPEEFWQLLKDGRDGISTFPTDRGWDLDLLAGDGSGSSATAEGGFIDA
                     ASFDAGFFGISPREAVTMDPQQRILLESSWEALERAGIDPVSLRGTETGVFVGTSGID
                     YINTVMNSREDIEGHGSTGLIASILSGRVSYTFGLEGPAVTVDTACSSSLVSLHLAAQ
                     ALRVGECSLALAGGVTAISTPMSFVGFSRQGGLAPDGRCKAFSDTADGTAWSEGVGVL
                     VVERLSDARRNGHPVLAVVRGSAVNQDGASNGLTAPSGPSQQRVIRQALAAAGLSASD
                     VDVVEAHGTGTALGDPIEAQALLATYGRDRDPERPLLLGSVKSNLGHTQAAAGAAGLI
                     KMILAMRHGVLPRTLHVEAPSSHVDWSSGALQLLTEPQEWPKADRPWRAGVSSFAISG
                     TNAHIILEQPPAAEEPEDEEGPHTSPSAVPWLVSAKSEEALTAQLDRISALTDRSPLD
                     LGLSLATTRAHLEHRAVLLSEGGETVEVARGHAEHTAGKLAVLFSGQGAQRVGMGREL
                     YGRFPVFAEALDGVLAHFDAGLRDVMFGEAEGLDETGFTQPALFAIEVALFRLSESLG
                     VRPDFVAGHSIGEIAAAHVAGVFSLADACALVAARARLMQALPAGGAMVAVQAAESEV
                     AARLVDGVSIAAVNGPEAVVIAGEEREVLRIAEEFTGLGRKTRRLPVSHAFHSPLMEP
                     MLEDFRRVAEGLSYEAPRIALVSNLTGELATEDLVRSAEYWVRHVRETVRFADGVSAL
                     AAEGASVFLELGPDGVLTAMAQHTLDGTATTVPALRKDRSEETALLTALAQLHVAGVG
                     VDWSGIFAGTGARRVDLPTYPFQHERYWPRPAVLSGDVVGAGLRSAEHPLLGATLTLA
                     ESGEVVLTGRLSLQAYPWLTDHTMGGSVLFPATGFLELAMRAADQVGCDRVDEFVLMT
                     PLVLAEESGTQVQVLVGPADEAGTRTISIHARPDGDPDQPWMQHAGGTLTSGERIVDF
                     GSTVWPPQGAEAVDLEGPYDTTGYGPAFQGLRSVWQCEDGAYVEVALPDQVAEDAGAF
                     GLHPVLLDAVLQSRRMAGIGAAGDTFLPFAWRGVSLHAGGASVLRARVIKTGEDTLSL
                     VAVDAEGAPVLSVEELVLRTQLTHTDAPLTARRGAERDALLSLEWVSGPEVPPAGVRC
                     VSLGADELGVGASVATLADLSGDEDFVVVPVSGVGGDVPAAAHELTARTLGLLQEWSG
                     REGTSASRLVFVTRNAVSADGGESVRDVAAAAVWGLVRSAQSESPGRIALVDLEADSG
                     DLAEVLAGLPGLLAAGDAQFVVREGAVRVGRLVESNEGAGVSPREWDADGTVLVTGGT
                     GGLGRELARHLVVERGVRHLLLVSRSGPEASGVGELRDELAAHGAEIAVRACDVADRA
                     AVEAVVGAVDAAHPLTAVIHTAGVLDDGVIASLTPERLSAVLRPKVDAAWNLHEATKD
                     LDLAAFVLYSSISGVLGSLGQGNYAAGNVFLDALAGHRAGLGLPAQSLAWGAWTQGGG
                     MTATLSDADMQRMASYGTAPLSPERGLALFDAATAGSATHLVAVGRVSGPVRAQGPVP
                     SMLRGLVKGARRTAVNTTGETAVSLADRLRAARPEERTQLMTDVVRTEAAAVLGHASS
                     QNVDARREFYELGFDSLTSVELRNRLSTVTGLRLSATVVFDSKNPTDLAARLYADLAA
                     QSALDGSASQAGFLPVAAPETDSLERMFLDALDSGKVPEAQRMLSALGALRPSFENTA
                     ELEDLPLPATLAEGPGTPRLICVSTPTANGGVHEYARLAASFRGERHVSALPLVGFAT
                     GERLPATAETAVRVIAESALRASDGNPPVLVGHSSAGAFAYLAAALLENTWGIRPEAV
                     VLLDTLSLRHEQNESIDYAGLMRRHFMVDEVSPVRMTNSRLSAMARWMGMLNQLEVRH
                     TTVPVLIIRATKETFGIGTDTGIYGEDHGSPVDVRSVDADHFSMVRDDAPETARIVRE
                     WLDSLGNA"
gene                 complement(9772. .15198)
                     /gene="pimS3"
CDS                  complement(9772. .15198)
                     /gene="pimS3"
                     /function="polyketide synthase"
                     /codon_start=1
                     /transl_table=11
                     /product="PimS3 protein"
                     /protein_id="CAC20920.1"
                     /db_xref="GI:12055071"
                     /translation="MTTSSETVVAALRASLKEAENLRRQNRKLVAAATEPIAIVAMAC
                     RYPGGVQSPEDLWELLTAGKDAIGAFPTDRGWDLDALRDAGVDRRGNAVSQEGGFLDG
                     VGDFDADFFGISPRESVTMDPQQRLLLETSWEAIERAGIDPTTLRGSRTGVFIGTNGQ
                     DYAYLLVRSLADATGDIGTGIAASATSGRLSYTLGLEGPAVTVDTACSSSLVALHTAS
                     HALRAGECSLALVGGVNVMSSPGSLMEFSRQGGLAADGRCKAFSDTADGTGWSEGVGV
                     LVVERLSDARRNGHPVLAVVRGSAVNQDGASNGFTAPSGPSQQRVIRQALAAAGLSAS
                     DVDVVEAHGTGTPLGDPIEAQALLATYGQDRDPERPLLLGSVKSNLGHTQAAAGAAGL
                     IKMILAMRHGVLPKTLYADVPSSHVDWTSGEVELLTEPQEWPKADRPWRAGVSSFGIS
                     GTNAHVIVEQAEPDEEPVAEPGLAPAVVPWLVSAKTEEALAAQLARISSRRGTASPVD
                     VGFSLASGRSAFEHRAVLLAQCPDGELSEAARGRAETSRPLAVLFSGQGAQRVGMGRE
                     LYVRFPVFAEALDEVLAHFDAGLRDVMFGDAEGLDETGFTQPALFAIEVALFRLAESL
                     GVRPDFVGGHSIGEIAAAHVAGVFSLTDACALVAARARLMQALPAGGAMVAVQASESE
                     VAARLVDGVSIAAVNGPEAVVLAGDETAVLRIAEELAAGGRKTQRLPVSHAFHSPLME
                     PMLEDFRRVAEGLSYGAPQIPLVSNVTGELAVEELVCSPDYWVRHVRETVRFADGVSA
                     LEAEGASAFLELGPDGVLTAMARHTLDETKVAVPVLRKDRPEETALLTALAHLHVAGI
                     GVDWSGIFAGTGARRVDLPTYPFQHERYWPRPAVLSGDVAGAGLRSAEHPLLGATLTL
                     AESGEVVLTGRLSLQTHPWLTDHAIGGNVFFPGTGFLELAMHAADQAGCDRVEDLVLM
                     TPLVLAEENGTQVQVLVGPADEAGTRTISIHARPEDDPDQPWMQHAGGTLTSGERVVD
                     FGSTVWPPQDAVAVDLEGFYEGTGYGPAFQGLRSVWSCGDGAYAEVALPDHVAEDAGA
                     FGLHPALLDAVLQSHRIAGVGSADDQFLPFAWRGVSLHAGGASVLRVRVVVKTGDDSMS
                     IAAVDAEGAPVLSAEAMVMRAYRDPGTGQAAARHGAERDELLSLEWVSGPEVPPAGVR
                     CVSLGADELGVGASVASLADLSGDEDFVVLPVSGVGGDVPAAAHELTARTLGLLQEWS
                     GREGSSGSRLVFVTRNAVSADEGESVRDVAAAAVWGLVRSAQSESPGRIALVDLEADS
                     GDLAEVLAGLPGLLAAGDAQFVVREGAVRVGRLVESNEGAGVSPREWDADGTVLVTGG
                     TGGLGRELARHLVVERGVRHLLLVSRSGPEAPGVSGLSDELAAHGAEVTVRACDVADR
                     AAVHALVAAVPAEHPLTAVIHTAGVLDDGLVTSLTPERLSTVLRPKVDAAWNLHEATK
                     DLDLAAFVLYSSISGVLGAPGQANYAAGNVFLDALAWHRMGQGLPAQSLAWGAWTQDG
                     GMTATLSDTDMQRMASYGTAPLSRQRGLALFDAACARSAAHLVAVGRVGGPVRASAPV
                     PSVLRGLVKGARRTAVNTTGETAVSLADRLRAARPEERPRLMTDLVRTEAAAVLGHAS
                     AKNVDARREFNELGFDSLTSVELRNRLSALTGLRLSAALVFDYPNPTALADHFVSELV
                     EEAATGPALLAELERLDSALAAGAPDARTRAAVTARLAQILDDWRDAPAEEDGAEVAE
                     RLEAASTDEIFAFIDNELGRHSDR"
gene                 complement(15210. .43733)
                     /gene="pimS2"
```

```
     CDS             complement(15210. .43733)
                     /gene="pimS2"
                     /function="polyketide synthase"
                     /codon_start=1
                     /transl_table=11
                     /product="PimS2 protein"
                     /protein_id="CAC20921.1"
                     /db_xref="GI:12055072"
                     /translation="MDQEQKLRDYLKRASADLRRSRQRVGELEAAAREPIAIVGMSCR
                     FPGGVRSPEELWSLLAADGDGITKSPADRGWELLAEDAGEFSGGFLHDAAEFDASFFG
                     ISPREALAMDPQQRLLLESSWEAFERAGIDPAAVRGTRTGMFVGAMPQEYRVGPDDDV
                     QGFALTGTTTSVISGRLAYFFGAVGPAVTVDTACSSSLVALHLAAHSLRQGECSLALA
                     AGVTVMSSPTTFVEFNRQGGLSADGRCRSFADSADGTGWSEGVGVLVLERLSEARRNG
                     HEILAVIRGSAVNQDGASNGLTAPNGPSQRRVIEQALVSARLSADEVDVVEAHGTGTT
                     LGDPVEAQALLATYGQGRDDDRPLLLGSVKSNLGHTQAAAGMAGVIKMVLAMRHGVLP
                     RTLHVDAPSSHVDWTQGAVRLLTEHVPWPQGEQPRRAGVSSFGLSGTNAHTILEEAPA
                     APSPEAGQEDTAPVAEPLAAGAVPWLVSGRTREALRAQAARLLEHLTSRPGLAPLDVA
                     YSLATARSGLEHRAAFTAAGPDGARAALTALADGTPAPGLVQDTARTRSKLAFLFAGQ
                     GSQRPGMGRELAARFPVFATALDEVLGHFDEGLERPLKDVLFAAEGTPEAALLDQTGY
                     AQPALFAIEVALYRLAESFGITPDFLAGHSIGEIAAAHVAGVFTLADAAALVLARGRL
                     MQALPEGGAMVSLEATEEEVLPLLEDRPDRLSIAAVNGPAAVVIAGEADDVTAVAEHF
                     DALGRRTKRLRVSHAFHSPLMEPMLAEFRTAVSRLTPQAPVIPVVSGLTGTLATVEQL
                     TSPDYWLDHARHAVRFADALRLARRARRRHLPRMGPDGVLSAMAQSCCRGHGHRRRRR
                     LRHARRVRPGRPEAETLTAALAGLHTRGVAVQWEPYFQGTGARRADLPTYAFQRRRYW
                     PKSLPATGGDVRAAGLGAAHHPLLTAAVSVANSDGLLLTGRLSRRTHPWLADHAVRGT
                     VLLPGTAFLELAVRAGDEAGCGRVEELTLAAPLVLPEEGGVQVQVWVGSPDASGRRAV
                     SVHSRPDGPEELPWTQHAAGTLVAGEHHAGFDATVWPPADATPLDLDGFYARMADTGF
                     GYGPLFQGLRAAWRSGDDVYAEVALPDSDGAASTESSAAFGLHPALLDAALHAAAFVD
                     LGEDARGGLPFSWQDVTLHASGASTVRVKLTPDGDDAVALAVADTTGAPVASIGSLVL
                     RTLPDEQIDAAHSLVRDALFELRWSEARQTAHEAPASAAVLGPDLFGLERALAPTDLT
                     VTAVPGPDAAPHAPDVVLAPVAGTTEDDVVTSVHARTAHVLERLRTWLTDERYAASRL
                     VFVTRGALATDARQASDPVAAAVWGLVRAAQSEHPGRFGLLDLDPGTTEPATEPLLRA
                     LALSSDEPQTAVRGTAVLTARLARAQAPQAATEWDPEGTVVLTGATGGLGRVLARHLV
                     AERGVRHLLLVSRRGADADGAGELMAELAAHGAQVTLAACDLADRAAVRRLLSGVPAE
                     HPVTAVVHSAGSLDDGVIASLTPERISAVLRPKADAVWHLHEATRDLDLAAFVVFSSL
                     SGTVGAAGQGNYAAANAFLDALAQLRRAQGLPGLSLAWGPWAPTAGMTGGLTDDDLDR
                     LARMGTPALTEEQGIALFDAATATDAAVLLPTRIDLSVLRVQSEIPPLWRGLIRTPAR
                     RSVVSASETAVTLVQQLSRLAETDRREVVLDLVAGQVAAVLGHAGTADIDPRRPLREL
                     GFDSLTAVELRNRLSAATGLRTAATVIFDHPTVDALAAHLLDGLMGSEAAAAADAPRT

  Query Match           55.6%;  Score 716.6;  DB 1;  Length 84985;
  Best Local Similarity  73.0%;  Pred. No. 3.7e-80;
  Matches  936;  Conservative    0;  Mismatches  344;  Indels    3;  Gaps    1;

Qy          1 GAGCCGATCGCCATCGTTGGCATGGCCTGTCGTTTCCCCGGCGGAGTGACCTCGGCGGAC 60
              |||||||||||| ||||| |  |||||||| || | |||||||| ||     ||| | ||
Db      15096 GAGCCGATCGCGATCGTGGCGATGGCCTGCCGCTACCCCGGCGGTGTCCAGTCGCCCGAG 15037

Qy         61 GACTTCTGGGATCTGATCTCCTCCGAGCAGGACGCGATCGGCGGATTCCCCACCGACCGC 120
              ||| | ||||| ||| |  || | |    ||||||| |||||||  |||||||| ||||||
Db      15036 GACCTGTGGGAGCTGCTGACCGCGGGCAAGGACGCCATCGGCGCGTTCCCCACGGACCGC 14977

Qy        121 GGCTGGGACCTGGACACGCTCTACGACCCCGACCCCGACCACCCCGGCACCTGCTACACC 180
              |||||||||||||||| | ||     ||| |||     |||||  | |||||  |  |  || |
Db      14976 GGCTGGGACCTGGACGCCCTGCGGGACGCCGGGGTCGACCGGCGCGGCAACGCCGTCAGC 14917

Qy        181 CGAAACGGCGGATTCCTCTACGACGCAGGCCACTTCGACGCCGAATTCTTCGGCATCAGC 240
              |   | || || ||||||  ||| ||  |||  |||||||||| || ||||||| |||
Db      14916 CAGGAAGGGGGCTTCCTCGACGGCGTGGGCGACTTCGACGCGGACTTCTTCGGGATCTCA 14857

Qy        241 CCCCGCGAAGCCCTCGCCATGGACCCCCAGCAACGACTCCTCCTCGAAACCGCCTGGGAA 300
              || |||||| || || |||||||||| ||||| || ||||| || || ||| |||||||
Db      14856 CCTCGCGAATCCGTCACCATGGACCCGCAGCAGCGCCTCCTGCTGGAGACCTCCTGGGAG 14797

Qy        301 ACCATCGAACACGCCGGCATCAACCCCCACACCCTCCACGGCACCCCCACCGGAGTCTTC 360
               ||||||||| |||||||||| ||||   ||||||  | |||||  || ||||||  || |||
Db      14796 GCCATCGAACGCGCCGGCATCGACCCGACCACCCTGCGCGGCAGCCGCACCGGCGTGTTC 14737

Qy        361 ACCGGCACCAACGGACAGGACTACGCACTTCGCGTGCACAACGCGGGCCAGTCAACCGAT 420
              | ||||||||||| ||||||||||    |   |     | |                ||
Db      14736 ATCGGCACCAACGGCCAGGACTACGCCTACCTGCTCGTACGCTCCCTGGCCGACGCCACC 14677

Qy        421 GGTTTCGCACTGACCGGAACCGCCGGCAGCGTCATCTCCGGTCGTATCTCGTACACGTTT 480
```

```
              ||   |     ||||||| ||||| ||||| || ||||||  |  | |||||||||  |
Db      14676 GGCGACATCGGCACCGGAATCGCCGCCAGCGCCACCTCCGGGAGGCTGTCGTACACGCTG 14617

Qy        481 GGTTTTGAGGGTCCTGCGGTGTCGGTGGACACGGCTTGTTCCTCGTCGTTGGTGGCTTTG 540
              ||  | ||||| || |||||  | || |||||||| || || || ||| | |||||  ||
Db      14616 GGCCTGGAGGGCCCGGCGGTCACCGTCGACACGGCGTGCTCGTCCTCGCTCGTGGCCCTG 14557

Qy        541 CATCTGGCCTGTCAGGCGTTGCGTGCGGGTGAGTGCTCGATGGCGCTTGCCGGGGGTGTG 600
              ||   |||  | || ||  |||| || ||||||||||||| |||| || |  ||| || |||
Db      14556 CACACGGCGAGCCACGCCCTGCGGGCCGGTGAGTGCTCGCTGGCCCTGGTCGGCGGGGTG 14497

Qy        601 ACGGTGATGTCGTCTCCGGGTGCCTTCGTGGAGTTTTCGCGGCAGCGGGGTCTGGCCGCG 660
              |  || |||||||| || ||| || |  |||||||      |||||  | || || || |||
Db      14496 AACGTCATGTCGTCCCCCGGTTCCCTGATGGAGTTCAGCAGGCAGGGCGGGCTCGCGGCG 14437

Qy        661 GACGGGCATTGCAAGGCGTTCTCGGCGGCGGCGGACGGGACCGGCTGGGGTGAGGGTGTG 720
              ||||| |  |||||||||||||| |   |||| |||||  |||||  |||    ||||| |||
Db      14436 GACGGCCGGTGCAAGGCGTTCTCCGACACGGCCGACGGCACCGGGTGGTCGGAGGGCGTG 14377

Qy        721 GGGATGCTGCTGGTGGAGCGGCTCTCCGACGCCCATCGCAACGGTCACCGTGTCCTGGCC 780
              ||  | ||| | |||||||| || ||||||||||  |||||||| || |  |||||  ||
Db      14376 GGCGTTCTGGTCGTGGAGCGCCTGTCCGACGCCCGCCGCAACGGGCATCCCGTCCTCGCG 14317

Qy        781 GTGGTGCGTGGCAGTGCGGTCAACCAGGACGGTGCGAGCAACGGTCTGACCGCGCCCAAC 840
              || || || |||  || |||||||||||| ||   |||||  | ||||||||||| |  |
Db      14316 GTCGTACGGGGCTCGGCCGTCAACCAGGACGGCGCCTCCAACGGCTTCACCGCGCCCAGC 14257

Qy        841 GGGCCGTCCCAGCAGCGTGTCATCCGCCAGGCCCTCGCCAACGCCGGCTTGTCGGCCGGT 900
              || || |||||||||||| || ||||| ||||| || |||  ||||||  ||||||||
Db      14256 GGTCCCTCCCAGCAGCGGGTGATCCGGCAGGCGCTGGCCGCCGCCGGTCTGTCGGCCTCG 14197

Qy        901 GATGTCGACGCGGTGGAGGCCCACGGCACCGGCACCACTTTGGGCGACCCGATCGAGGCC 960
              || || |||| ||| ||||||||||||||||||||| |  | || || ||||||||||
Db      14196 GACGTGGACGTGGTCGAGGCCCACGGCACCGGCACCCCGCTCGGTGATCCGATCGAGGCG 14137

Qy        961 CAGGCCCTCCTCGCGACCTACGGACAGGACCGTGCCGGCGAGGGGCCGCTGTGGCTGGGC 1020
              ||||| || || || || ||||| |||||||| | |  |||| ||||||||  ||||||
Db      14136 CAGGCGCTGCTGGCCACGTACGGTCAGGACCGCGACCCCGAGCGGCCGCTGCTGCTGGGG 14077

Qy       1021 TCGGTCAAGTCCAATGTCGGTCACACACAGGCTGCCGCGGGCGTCGCCGGGGTGATCAAG 1080
              |||||||||||||  |||| || || ||||| || || ||||  || ||| ||||||||
Db      14076 TCGGTCAAGTCCAACCTCGGCCATACCCAGGCGGCGGCCGGCGCGGCAGGGCTGATCAAG 14017

Qy       1081 ATGGTGATGGCGCTGCGGCATGGTCTGCTGCCGCGGACGTTGCATGTGGATGAGCCGTCG 1140
              ||| |  ||||| |||||||  ||| |  |||||    |||| || | |  |||| ||||||
Db      14016 ATGATCCTGGCGATGCGGCACGGTGTCCTGCCCAAGACGCTGTACGCCGATGTGCCGTCC 13957

Qy       1141 CCGCATGTGGACTGGTCCGCGGGTGCGGTGCAGCTGCTGACGGAGACGGTGCCCTGGCCC 1200
              | || ||||||||  |  |||| | |||| ||||||||||||| |   |  |||||
Db      13956 TCCCACGTGGACTGGACGTCGGGGGAGGTGGAGCTGCTGACGGAGCCTCAGGAGTGGCC- 13898

Qy       1201 GGCGGGGAGGGGCGGCTACGGCGGGCAGGAGTGTCATCATTCGGCGTCAGCGGCACCAAC 1260
                |  ||  |  ||  |  |||||||  || |||||  || ||||||  ||||||||||||||
Db      13897 --CAAGGCCGACCGCCCGTGGCGGGCCGGCGTGTCCTCGTTCGGCATCAGCGGCACCAAC 13840

Qy       1261 GCCCACGTCATCCTCGAAGAAGC 1283
              ||||| || ||| | ||  | ||
Db      13839 GCCCATGTGATCGTGGAGCAGGC 13817
```

```
RESULT 9
AX211705
LOCUS       AX211705               65140 bp    DNA     linear   PAT 06-SEP-2001
DEFINITION  Sequence 1 from Patent WO0159126.
ACCESSION   AX211705
VERSION     AX211705.1  GI:15523937
KEYWORDS    .
SOURCE      Streptomyces noursei.
  ORGANISM  Streptomyces noursei
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1  (bases 1 to 65140)
  AUTHORS   Zotchev,S.B., Sekurova,O.N., Fjaervik,E., Brautaset,T.,
```

```
            Stroem,A.R., Valla,S., Ellingsen,T.E., Sletta,H.V. and
            Gulliksen,O.M.
  TITLE     Gene cluster encoding a nystatin polyketide synthase and its
            manipulation and utility
  JOURNAL   Patent: WO 0159126-A 1 16-AUG-2001;
            Norges Teknisk Naturvitenskapelige Universitet (NO) ; STIFTELSEN
            IND OG TEKNISK FORSKNING VED NORGES TEKNISKE HOGSKOLE (NO) ;
            ALPHARMA AS (NO) ; Sinvent AS (NO) ; Zotchev, Sergey Borisovich
            (NO) ; Sekurova, Olga Nikolayivna (NO) ; Fjaervik, Espen (NO) ;
            Brautaset, Trygve (NO) ; Strom, Arne Reidar (NO) ; Valla, Svein
            (NO)
FEATURES             Location/Qualifiers
     source          1. .65140
                     /organism="Streptomyces noursei"
                     /db_xref="taxon:1971"
                     /note="ATCC 11455"
BASE COUNT     8270 a  25171 c  22273 g   9426 t
ORIGIN

  Query Match           55.5%;  Score 716;  DB 6;  Length 65140;
  Best Local Similarity   73.3%;  Pred. No. 4.7e-80;
  Matches  946;  Conservative    0;  Mismatches  335;  Indels    9;  Gaps    2;

Qy          1 GAGCCGATCGCCATCGTTGGCATGGCCTGTCGTTTCCCCGGCGGAGTGACCTCGGCGGAC 60
              || ||||||| |||||| |||||||||| || | |||||||||  |  ||||  | ||
Db      21818 GACCCGATCGTCATCGTCGGCATGGCCTGCCGCTACCCCGGCGGCATCGCCTCACCCGAG 21877

Qy         61 GACTTCTGGGATCTGATCTCCTCCGAGCAGGACGCGATCGGCGGATTCCCCACCGACCGC 120
              ||| |||||   ||| ||  |   |     ||||| |  |||   |||||||| |||||
Db      21878 GACCTCTGGCGCCTGGTCAGCCAGGGCGCCGACGCCACTGGCCCGTTCCCCACCAACCGC 21937

Qy        121 GGCTGGGACCTGGACACGCTCTACGACCCCGACCCCGACCACCCCGGCACCTGCTACACC 180
              |||||||||||||||  |||||||||||||||||||||| ||| |||  |  | ||  |
Db      21938 GGCTGGGACCTGGACAACCTCTACGACCCCGACCCCGACCGCCCGGGCCGCACCCACGTC 21997

Qy        181 CGAAACGGCGGATTCCTCTACGACGCAGGCCACTTCGACGCCGAATTCTTCGGCATCAGC 240
              ||   |||||| |||||  ||||||| |||  |||||||||||| |||||||| || |||
Db      21998 CGCGCCGGCGGCTTCCTGCACGACGCCGGCTCCTTCGACGCCGACTTCTTCGGGATGAGC 22057

Qy        241 CCCCGCGAAGCCCTCGCCATGGACCCCCAGCAACGACTCCTCCTCGAAACCGCCTGGGAA 300
              || ||||| ||  | ||||  ||| ||||||| || || || ||||||  | ||||||||
Db      22058 CCGCGCGAGGCGATGGCCACCGACTCCCAGCAGCGCCTGCTGCTCGAACTCTCCTGGGAA 22117

Qy        301 ACCATCGAACACGCCGGCATCAACCCCCACACCCTCCACGGCACCCCCACCGGAGTCTTC 360
              || |||||| |||||||||| |||||  | | || | || | ||  ||||| ||||||
Db      22118 GCCGTCGAACGCGCCGGCATCGACCCCGCCTCACTGCGCGACTCCGGCACCGGCGTCTTC 22177

Qy        361 ACCGGCACCAACGGACAGGACTACGCACTTCGCGTGCACAACGCGGGCCAGTCAACCGAT 420
              |||||  ||      | |||||||      ||     |   | ||| |     |||
Db      22178 GCCGGCGTCATGTACAACGACTACG------GCACCACCCTGACCGGCGACGAGTACGAG 22231

Qy        421 GGTTTCGCACTGACCGGAACCGCCGGCAGCGTCATCTCCGGTCGTATCTCGTACACGTTT 480
              |  |||       | ||| | ||||   ||||||  ||||||  ||  |||| |||||  |
Db      22232 GCGTTCCGCGGCAACGGCAGCGCCCCGAGCGTCGCCTCCGGCCGCGTCTCCTACACCCTC 22291

Qy        481 GGTTTTGAGGGTCCTGCGGTGTCGGTGGACACGGCTTGTTCCTCGTCGTTGGTGGCTTTG 540
              ||  | ||  ||  ||  ||  ||  |||||||||| ||  ||  ||  ||  ||  |||| ||  ||
Db      22292 GGCCTGGAAGGCCCGGCCGTCACGGTGGACACCGCCTGCTCTTCCTCTCTGGTCGCCCTG 22351

Qy        541 CATCTGGCCTGTCAGGCGTTGCGTGCGGGTGAGTGCTCGATGGCGCTTGCCGGGGGTGTG 600
              ||   |||    |||||||||| ||||| ||||||||| |||||  | |||||  ||||||
Db      22352 CACTGGGCGGCGCAGGCGTTGCGGGCGGGGGAGTGCTCGTTGGCGTTGGCCGGTGGTGTG 22411

Qy        601 ACGGTGATGTCGTCTCCGGGTGCCTTCGTGGAGTTTTCGCGGCAGCGGGGTCTGGCCGCG 660
              |||||||||||  | |||  |  | ||||||||||| ||||||||||||||||||  |
Db      22412 ACGGTGATGTCGACGCCGAGCACGTTCGTGGAGTTCTCGCGGCAGCGGGGTCTGGCGCCT 22471

Qy        661 GACGGGCATTGCAAGGCGTTCTCGGCGGCGGCGGACGGGACCGGCTGGGGTGAGGGTGTG 720
              ||  ||  |  ||  |||||||||  |  |  |||  ||||||||     | ||||    |||||  ||
Db      22472 GATGGTCGTTCGAAGGCGTTCGCCGAGGCCGCGGACGGCGTGGCCTGGTCCGAGGGCGTC 22531

Qy        721 GGGATGCTGCTGGTGGAGCGGCTCTCCGACGCCCATCGCAACGGTCACCGTGTCCTGGCC 780
              ||  ||||||  |  |||||||||  ||  |||||    ||||||||||||    ||||||||

Db      22532 GGCATGCTGGTCCTGGAGCGGCAGTCGGACGCGGTGCGCAACGGTCACGAGATCCTGGCC 22591

Qy        781 GTGGTGCGTGGCAGTGCGGTCAACCAGGACGGTGCGAGCAACGGTCTGACCGCGCCCAAC 840
              |||||||| |||   |||||||||||||||||||||  |||||||||||||||||||||
Db      22592 GTGGTGCGCGGCTCGGCGGTCAACCAGGACGGTGCGTCCAACGGTCTGACCGCGCCCAAC 22651

Qy        841 GGGCCGTCCCAGCAGCGTGTCATCCGCCAGGCCCTCGCCAACGCCGGCTTGTCGGCCGGT 900
              || |||||||||||||| || ||||| |||||  | ||||  | ||||  ||||  | |
Db      22652 GGCCCGTCCCAGCAGCGGGTGATCCGTCAGGCGTTGGCCAGTGGCGGCCTGTCCACGGCC 22711

Qy        901 GATGTCGACGCGGTGGAGGCCCACGGCACCGGCACCACTTTGGGCGACCCGATCGAGGCC 960
              || || |||| || ||||| |||||||| || || ||  | || ||||||||||||||||
Db      22712 GACGTGGACGCCGTTGAGGCGCACGGCACGGGTACGACGCTCGGTGACCCGATCGAGGCC 22771

Qy        961 CAGGCCCTCCTCGCGACCTACGGACAGGACCGTGCCGGCGAGGGGCCGCTGTGGCTGGGC 1020
              ||||| ||||| || |||||||| |  |||||  |  |  ||||   ||||||  ||| |||
Db      22772 CAGGCGCTCCTGGCCACCTACGGTCGCGACCGCGACCCCGAGAACCCGCTGCTGCTCGGC 22831

Qy       1021 TCGGTCAAGTCCAATGTCGGTCACACACAGGCTGCCGCGGGCGTCGCCGGGGTGATCAAG 1080
              ||| ||||||||||  |||||||||| ||||| || || || ||||||||  || ||||||
Db      22832 TCGATCAAGTCCAACATCGGTCACACCCAGGCAGCGGCCGGTGTCGCCGGTGTCATCAAG 22891

Qy       1081 ATGGTGATGGCGCTGCGGCATGGTCTGCTGCCGCGGACGTTGCATGTGGATGAGCCGTCG 1140
              ||||| |||||| ||||||| ||  |||||||| |||  ||||||| || | ||||||
Db      22892 ATGGTCATGGCGATGCGGCACGGCGTGCTGCCGCAGACCCTGCATGTCGACGCGCCGTCC 22951

Qy       1141 CCGCATGTGGACTGGTCCGCGGGTGCGGTGCAGCTGCTGACGGAGACGGTGCCCTGGCCC 1200
                |||| || || |||  ||  || || ||  | ||||| || |||  |    |||||||
Db      22952 TCGCACGTCGATTGGAGCGTCGGCGCCGTCGAACTGCTCACCGAGCAGACCGCCTGGCC- 23010

Qy       1201 GGCGGGGAGGGGCGGCTACGGCGGGCAGGAGTGTCATCATTCGGCGTCAGCGGCACCAAC 1260
                | |   || |||   || || || || || || || || |||||| ||||||||||||||
Db      23011 --GGAGACCGGCCGGGCCCGTCGCGCCGGTGTCTCCTCCTTCGGCATCAGCGGCACCAAC 23068

Qy       1261 GCCCACGTCATCCTCGAAGAAGCACCCGCC 1290
              ||||||||| || ||||  | | ||  ||
Db      23069 GCCCACGTCGTCATCGAGCAGTCCCCGACC 23098


RESULT 10
AF263912
LOCUS       AF263912               123580 bp    DNA     linear   BCT 24-MAY-2000
DEFINITION  Streptomyces noursei ATCC 11455 nystatin biosynthetic gene cluster,
            complete sequence.
ACCESSION   AF263912
VERSION     AF263912.1  GI:8050835
KEYWORDS    .
SOURCE      Streptomyces noursei.
  ORGANISM  Streptomyces noursei
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1  (bases 1 to 123580)
  AUTHORS   Brautaset,T., Sekurova,O.N., Sletta,H., Ellingsen,T.E., StrLm,A.R.,
            Valla,S. and Zotchev,S.B.
  TITLE     Biosynthesis of the polyene antifungal antibiotic nystatin in
            Streptomyces noursei ATCC 11455: analysis of the gene cluster and
            deduction of the biosynthetic pathway
  JOURNAL   Chem. Biol. 7 (6), 395-403 (2000)
  MEDLINE   20334850
   PUBMED   10873841
REFERENCE   2  (bases 1 to 123580)
  AUTHORS   Brautaset,T., Sekurova,O.N., Sletta,H., Ellingsen,T.E., Strom,A.R.,
            Valla,S. and Zotchev,S.B.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-MAY-2000) Unigen, NTNU, O. Kyrres gt. 3, Trondheim
            N-7489, Norway
FEATURES             Location/Qualifiers
     source          1. .123580
                     /organism="Streptomyces noursei"
                     /strain="ATCC 11455"
                     /db_xref="ATCC:11455"
                     /db_xref="taxon:1971"
     gene            complement(46. .783)
```

```
                /gene="nysF"
     CDS        complement(46. .783)
                /gene="nysF"
                /function="presumably post-translationally modifies ACP
                domains on PKS"
                /note="putative 4'-phosphopantheteine transferase"
                /codon_start=1
                /transl_table=11
                /product="NysF"
                /protein_id="AAF71762.1"
                /db_xref="GI:8050836"
                /translation="MIELILPATVATEAAYDDRPRPGDRLLSSEREVIARAVESRQRE
                FTTVRHLARRALRRLGHPDRAILPNRRGAPQWPPGIVGSMTHCAGYRAAAVSPAELSA
                AVSIDAEPNGPLPAGVLNAIALPSERPHLVALAAHRPDVHWDRLLFSAKESVFKAWYP
                LTQRELDFSEAEIVIDPTQGAFTARLLVPGPLLGGRRVTVFPGRWHSTPALLTTAVHL
                PAPTPRRDREHRTHLTVNSPLPRPTFG"
     gene       complement(867. .2684)
                /gene="nysG"
     CDS        complement(867. .2684)
                /gene="nysG"
                /function="presumably involved in efflux of nystatin"
                /note="putative transporter (ABC family)"
                /codon_start=1
                /transl_table=11
                /product="NysG"
                /protein_id="AAF71763.1"
                /db_xref="GI:8050837"
                /translation="MASPDDLEEERTAPRPARRLVGLLRPHRRSVALAVSMGVGGIVL
                NAFGPLLLGRVTDLIADGVLGGVPGPAPGIDFAAIGRLLLVLLALYVVASLFMLAQGR
                LVASAVWRTIHELRRDAREKLTRLPLRHFDRQPAGELLSRTTNDIDNLQQTLQQTLAE
                LITSIFSLLTMLVLMLVISPSLAVVMLLSVPVSALIAARISKRAQPHYAAQWSANGTL
                NAHVEEVCTGHALIKGFDRRAAAEERFDACNDAVYRAAAKAQFASGAMEPVMMFVANL
                GYVLVAVIGAWKVINGTLTLGDVQAFILYARQFSQPIVEIASVAGRLQSGIASAQRVF
                TLLDAPEQAPDPLRPGTPARAEGRVEFTDVSFRYSPDTPLIENLSLTVEPGSTVAIVG
                PTGAGKTTLGNLLMRFYEPDSGRILLDGTDTATMTRDDLRSRFGLVLQDTWLFGGTIA
                ENIAYGAPGACRADIEEAARATCADRFIRTLPQGYDTVLDDESGTVSAGEKQLLTVAR
                AFLARPAVLVLDEATSSVDTRTEVLIQRAMNSLRAGRTSFVIAHRLSTIRDADLIVVM
                DAGRIVEQGTHDQLLCAQGLYARLHAARTHTPTAGAAAG"
     gene       complement(2662. .4416)
                /gene="nysH"
     CDS        complement(2662. .4416)
                /gene="nysH"
                /function="presumably involved in efflux of nystatin"
                /note="putative transporter (ABC family)"
                /codon_start=1
                /transl_table=11
                /product="NysH"
                /protein_id="AAF71764.1"
                /db_xref="GI:8050838"
                /translation="MLLRLLRAQLRPYAWATAALVALQLVQILGTLLLPTLGAALIDQ
                GVVRGDGGRITELGVVMGVVALVQIAAALGAAALAARTATAMGRDLRSALFRRILDFS
                AREIGRFGTPSLLTRSVNDVQQVQNLAQTGFGIVVCAPLMCLGSVLLALRQDVPLALL
                LVALVLVVAVCFGLLLARMGTLYARMQLTLDRLGRLLREAITGVRVVRSFVRDDHERA
                RFAQTNDAFLVVSRRVGRLIATMLPVVLLLMNGFTVALLWTGSHRIDAGRMPIGSLSA
                LLSYLSLILMSVVMLAFVFLSVPRARVCAGRIAEVLDTGSSVAPPAAPQPVRGPAGRI
                ELCAAGYRYPGAEEPVLRDVDLTVEPGERIAVLGSTGSGKTTLLNLVLRLADATEGAV
                RVGGTDVRELTAATLAAAVGFVPQRPYLFSGTVASNLRFGRPDATDEELWEALRVAQA
                ADFVARMPDGLDAEITQGGGNVSGGQRQRLSLARALLRRPEIYLFDDCFSALDQATDA
                ALRTALVPYTAGATVITVAQRISAGRDADRIVVLDRGRVVAQGTHDVLLRTSPTYREI
                ALSQLTEEEAAHGLAGRP"
     gene       4714. .5748
                /gene="nysDIII"
     CDS        4714. .5748
                /gene="nysDIII"
                /function="presumably involved in mycosamine biosynthesis"
                /note="putative dGDP-mannose-4,6-dehydratase"
                /codon_start=1
                /transl_table=11
                /product="NysDIII"
                /protein_id="AAF71765.1"
                /db_xref="GI:8050839"
                /translation="MSKRALITGITGQDGSYLAEHLLSQGYQVWGLIRGQANPRKSRV
                SRLASELDFIDGDLMDQGSLVSAVDTVQPDEVYNLGAISFVPMSWQQAELVTEVNGMG
                VLRMLEAIRMVSGLSTSRTVSPRGQIRFYQASSSEMFGKAAETPQRETTLFHPRSPYG
                AAKAYGHYITRNYRESFGMYAVSGMLFNHESPRRGQEFVTRKISLAVARIKQGLQDKL
                ALGNLDAVRDWGYAGDYVRAMHLMLQQDAGDDYVIGTGQMHSVRDAVRIAFEHVGLNW
                EDYVVIDPDLVRPAEVEVLCADSAKAQDRLGWKPDVDFPTLMRMMVDSDLAQVSRENQ
                YGDVLLAANW"
     gene       5930. .34363
                /gene="nysI"
     CDS        5930. .34363
                /gene="nysI"
                /function="responsible for condensation steps 9 to 14 in
                the nystatin polyketide backbone synthesis"
                /note="polyketide synthase"
                /codon_start=1
                /transl_table=11
                /product="NysI"
                /protein_id="AAF71766.1"
                /db_xref="GI:8050840"
                /translation="MDNEQKLRDYLKLATADLRRTRRRVHKLESAAQEPVAIIGMTCR
                YPGGVRSPEDLWRMVEAGEHGVTPFPTDRGWDLEALAAAPTASGGFLHDAPDFDADFF
                GISPREAVAMDPQQRVVLESAWEAFERAGIDPTSVKGSRTGVFIGAMAQDYRVGPADG
                AEGFQLTGNTGSVLSGRISYTFGTVGPAVTVDTACSSSLVAVHLATQALRAGECTLAL
                AGGVTIMSGPGTFIEMGRQGGLSADGRCRSFGDTADGTGWAEGVGILVLERLSDAVRN
                GHEILAVVRGTAVNQDGASNGLTAPNGPSQQQVIQQALVNARLAAGDIDVVEAHGTGT
                TLGDPVEAQALLATYGQNRPADRPLLLGSVKSNLSHTQAAAGVAGVIKMVMAMRHGTL
                PRTLHAEEPTHHVDWSQGAVRLLTDTTDWPATGAPRRAAVSSFGISGTNAHTIIEQAP
                EPQPEDAATAQDDAAGSTPATAPVVPGVVPVLLSGRTPDALRGQAAALRAALDTGRRP
                DLLDLAHSLATTRAGFEHRAVLLATDHPALTDGLTALADADDPAAAPAWITGTTRAET
                RLAVLFTGQGAQRLGAGRELAARFPAFATALDAALDAFTPHLDRPLREVLWGTDAALL
                DRTAYAQPALFAVEVALYRLIESFGVRPDHLAGHSVGEIVAAHLAGVLSLADAATLVA
                ARGRLMQALPDGGAMIAVQASEADVAPLLAGHEDQVAIAAVNGPSAVVLSGAEATVTA
                LAEQLAADGRKTRRLRVSHAFHSPLMEPMLDAFRAVVEDLTLQPPLLPVVSNLTGKPA
                TVAQLTSADYWVDHVRHAVRFADGIDWLARHDTTAFLELGPDGVLSAMAQDCLDAADA
                DAVTLPALRAGRPEEHTLTTALAGLHVHGATLDWTGCFAGTGARRTDLPTYAFQRRRY
                WPKALQSGTADLRSVGLGAAHHPLLSAAVSLADAGGTLLTGRLSRQTHPWLADHTVRG
                TTLLPGTAFLELAVRAGDEVGCDRVEELTLAAPLLLPEQGGVQVQLWIGNPDVSGRRT
                VNVHARPDTGDDTPWTAHATGVLTTADASRQLPASSEQGGTPLAGDPHPALDAAQWPP
                AGAEPLPLDGHYDRLADGGFGYGPVFQGLRAAWRGGDVVYAEVELPEAGRSDAEAFGL
                HPALLDAALHAAPFTGLGERGRGGLPFSWEGVSLHAGGATTLRVRLTPVADDALALTV
                ADGTGAPVLSVDSLVLRSVATQQLDTAAAVARDALFRLDWTPVQPTATDPGPVALLGA
                DPFGLLTHAGFADAPAYPDLAALAAADGPVPTTVVLSLAGTGDDAADPARSAHRCAAE
                ALAAVQTWLDHHERFAAARLVFVTRGATVGRDVAAAAVWGLVRSAQSENPGCFALVDL
                DPDGAVGAAALVAALVSGEPQLAVRGDVLRVARLVRRPLTEVGAGADGTGDGVGDGSG
                VSFSGEGAVLVTGGTGGLGAVLARHLVAEYGVRDLLLVSRSGERAVGAGELVAELAGV
                GARVRVVACDVTDRAAVVELVGGHAVSAVVHAAGVLDDGMVGALTGERLSAVLRPKVD
                AVWHLHEATRGLDLDAFVVFSSLAGVFGSPGQANYAAANAFLDALMTRRRAEGLPGLS
                LAWGPWEQSGGMTGTLTDVDAERLARSGVPPLSVAQGLALFDAAVAGTDATCVPVRLD
                LPVLRARGEVPPLLRSLIRVRARRAAVAGSATAGNLAQRLRRLDEDGRDEMVLDLVRG
                QVALVLGHATGGDVDAGRAFRDLGFDSLTAVELRNRLNTVTGLRLPATLVFDYPTVRH
                LATYVLDELLGTDAEVATVQPAAVAVADDPIVIVGMACRYPGGVSSPEDLWRVLTEGT
                DAVSGFPTNRGWDVESLYHPDPDHPGTSYTRSGGFLHEAGEFDPGFFGMSPREALATD
                SQQRLLLESSWEAIERAGIDPVSLRGSRTGVFAGVMYSDYSAMLASPEFEGFQGSGSS
                PSLASGRVAYTLGLEGPAVTVDTACSSSLVAMHWAMQALRSGECGLALAGGVTVMSTP
                AVFVDFARQRGLSPDGRCKAFADAADGVGWSEGVGVLVLERQSDAVRNGHEILAVVRG
                SAVNQDGASNGLTAPNGPSQQRVIRQALASGGLTAGDVDVVEAHGTGTTLGDPIEAQA
                LLATYGRDREPERPLLLGSVKSNLGHTQAAAGVAGVIKMVLAMRHGVVPRTLHVDAPS
                SHVDWSEGAVELLSEQAAWPETGRVRRAGVSSFGISGTNVHVIVEQAPGAKAIAAAGA
                ARRTPGAVPVLLSGRGRSALRGQAARLLGHLQARPDAELVDVALSLATTRSRFEQRAA
                VVAQDRDQLIASLGALAADRPDPAVVEGEAAGRGRTAVLFTGQGSQRAAMGRELHEVQ
                PEFAAAFDAVCAVFDPLLDRPLREVVFAEDGSDEAALLDETGWTQPALFAVEVALFRL
                VESWGVRPDFVAGHSIGEIAAAHVAGVLTLEDACRLVAARATLMQALPTGGAMIAIQA
                TEDEIAAHLDDTVAIAAVNGPQSVVISGDEEAAETIAATFAERGRKTKRLRVSHAFHS
                PRMDGMLDAFRIVAEGLTYRAPRIPLVSDLTGRRADDAEVCTAEYWVRHVREAVRFAD
                CVRTLRDAGATTFLELGSDGLLTAMAEDTLGDDHDAELVPMLRAGRAEELAAATALAR
                LQVRGVDVDWAAYLAGTGARRTDLPTYAFQHAYYWPQLPTPAAALAAADPADQQLWAA
                VERGDARELADILGLGEQDLTPLDSLLPALTSWRRGNQEKHLLDTLRYRVEWTRLSKP
                TAPVLDGTWLLVASDATAADQPALLDGLADALGSHGARVRRLLLDDSCADRAVLAERL
                ARTADVDAATQVLSVLPLDERDADDCPPLTRGLALTVALVQALADTGAQGRLWTATRG
                AVSTNPADPVTHPVQAAAWGLGRGVALEHPRLWGGLVDLPQVFDERAGQRLAGILAVK
                DAPDGEDQVALRATGVSGRRLVRHTVEALPTAAEFTATGTVLITGGTGGLGAEVARWL
                ARAGAQHLVLTSRRGPDAPGAAELRAELEGYGPSVSVVACDVADRDALAAVLTALPEE
                LPLTGVVHTAGVGHYGPLDTLSTAEFAGLTAAKLAGAAHLDALLADRELDFFVLFGSI
                AGVWGSGNQSAYGAANAYLDALALHRRARGLAATSVAWGPWAEAGMAADDAVSETLRR
                QGLGLLDPAPAMTELRRAVVRQDVTVTVADVDWQRYAPLFTSARPSALIAGLPEVRAL
```

```
                     AADERTEQDATGASEVVTRVRALAEPEQLRLLTDLVRTESATVLGHSSADAVPEGRAF
                     RDVGFDSLTAVELRKRLGAATGLSLPSTMVFDYPTPLELAQYLRAEILGAVLEVAGPV
                     ATGGADDEPIAIIGMACRFPGGVSSPEQLWDLVASGTDAISEFPVNRGWQTGHLFDPD
                     PDRPGTTYSTQGGFLHEADEFDPTFFGISPREALVMDPQQRLLLETTWESFERAGIRP
                     ETLRSTLTGTFVGSSYQEYGLGAGDGTEGHMVTGSSPSVLSGRLSYVFGLEGPAVTVD
                     TACSSSLVALHLACQSLRNGESNLAVAGGATIMTTPNPFIAFSRQRALAKDGRCKAFS
                     DDADGMTLAEGVGVVLVERLSDAQRNGHPVLAVLRGSAINQDGASNGLTAPNGPSQQR
                     VIRQALANARLAPGDIDALEAHGTGTPLGDPIEAQALFATYGRDRDPESALLLGSVKS
                     NIGHTQSAAGIASVIKMVMALRHSELPPTLHADAPSSHVDWSAGTVRLLTQARAWPET
                     GRPRRAAVSSFGISGTNAHVLLEQAPVADTPAEERPAVAPVPIAAGVVPWVVTARSAA
                     ALRGQAERLLAHAETVGTALPAAGPLDIGLSLVSARARFEHRAVVVPPAGTDPLAALR
                     AVATDGPSPVVARGVADVEGRTVFVFPGQGSQWVGMGSQLLDESAVFAERIAECAAAL
                     AEFTDWSLVDVLRGVVGAPSLERVDVVQPASFAVMVSLAALWRSRGVLPDAVVGHSQG
                     EIAAAVVSGALSLRDGARVVALRSQAIGRALAGRGGMMSVALSVDVLEPRLVEFEGRV
                     SVAAVNGPRSVVVAGEPEALDALHARLTADDIRARRIAVDYASHSHQVEDLHEELLEV
                     LAELAPRTSEVPFFSTVTGDWLDTARMDAGYWFRNLRGRVRFADAVADLLAAEYRAFV
                     EVSSHPVLSMAVQEAIDEAGVPAVAAGTLRRDQGGTDRFLLSAAEVFVRGVDVDWAGL
                     FEGTGASRIDLPTYAFQHEHLWAVPPAPEAVAAADPDDAAFWTAVEDGDVSALTAALG
                     TDEDSVAAVLPALTSWRRARRDRSTVDAWRYRVAWKPLGGTLPHPSLTGTWLLVTADG
                     IDDTDVAGALETYGAEVRRLVLDEECVDRAVLRERLAGAEDVTGIVSVLAAAERTDAV
                     PGTSLVLGTALTVALIQALGDAEIDAPVWALTRGAVSTGRADELTAPVQAQVTGIGWT
                     AALEHPQRWGGTLDLPAALDARAAQRLAAVLSGALGSDDQLAIRPSGVFTRRIVRAEA

  Query Match             55.5%;  Score 716;  DB 1;  Length 123580;
  Best Local Similarity   73.3%;  Pred. No. 4e-80;
  Matches  946;  Conservative    0;  Mismatches  335;  Indels    9;  Gaps    2;

Qy          1 GAGCCGATCGCCATCGTTGGCATGGCCTGTCGTTTCCCCGGCGGAGTGACCTCGGCGGAC 60
              || ||||||| |||||| |||||||||| || | |||||||||   |  ||||  | ||
Db      81671 GACCCGATCGTCATCGTCGGCATGGCCTGCCGCTACCCCGGCGGCATCGCCTCACCCGAG 81730

Qy         61 GACTTCTGGGATCTGATCTCCTCCGAGCAGGACGCGATCGGCGGATTCCCCACCGACCGC 120
              ||| |||||   ||| ||  |   |    |||||  |  |||   |||||||| |||||
Db      81731 GACCTCTGGCGCCTGGTCAGCCAGGGCGCCGACGCCACTGGCCCGTTCCCCACCAACCGC 81790

Qy        121 GGCTGGGACCTGGACACGCTCTACGACCCCGACCCCGACCACCCCGGCACCTGCTACACC 180
              ||||||||||||||||  |||||||||||||||||||||| ||| |||  |  | ||  |
Db      81791 GGCTGGGACCTGGACAACCTCTACGACCCCGACCCCGACCGCCCGGGCCGCACCCACGTC 81850

Qy        181 CGAAACGGCGGATTCCTCTACGACGCAGGCCACTTCGACGCCGAATTCTTCGGCATCAGC 240
              ||   |||||| ||||  |||||| |||  |||||||||||| ||||||| || |||
Db      81851 CGCGCCGGCGGCTTCCTGCACGACGCCGGCTCCTTCGACGCCGACTTCTTCGGGATGAGC 81910

Qy        241 CCCCGCGAAGCCCTCGCCATGGACCCCCAGCAACGACTCCTCCTCGAAACCGCCTGGGAA 300
              || ||||| ||  | ||||  ||| ||||||| || || || ||||||  | ||||||||
Db      81911 CCGCGCGAGGCGATGGCCACCGACTCCCAGCAGCGCCTGCTGCTCGAACTCTCCTGGGAA 81970

Qy        301 ACCATCGAACACGCCGGCATCAACCCCCACACCCTCCACGGCACCCCCACCGGAGTCTTC 360
              || |||||| |||||||||| |||||  | | || | || | ||  ||||| ||||||
Db      81971 GCCGTCGAACGCGCCGGCATCGACCCCGCCTCACTGCGCGACTCCGGCACCGGCGTCTTC 82030

Qy        361 ACCGGCACCAACGGACAGGACTACGCACTTCGCGTGCACAACGCGGGCCAGTCAACCGAT 420
              |||||  ||      | |||||||      ||      |  ||| |       |||
Db      82031 GCCGGCGTCATGTACAACGACTACG------GCACCACCCTGACCGGCGACGAGTACGAG 82084

Qy        421 GGTTTCGCACTGACCGGAACCGCCGGCAGCGTCATCTCCGGTCGTATCTCGTACACGTTT 480
              |  |||      | ||| | ||||   ||||||  ||||||  ||  |||| |||||  |
Db      82085 GCGTTCCGCGGCAACGGCAGCGCCCCGAGCGTCGCCTCCGGCCGCGTCTCCTACACCCTC 82144

Qy        481 GGTTTTGAGGGTCCTGCGGTGTCGGTGGACACGGCTTGTTCCTCGTCGTTGGTGGCTTTG 540
              ||  | || || || || ||  |||||||||| || || || || ||  |||| ||  ||
Db      82145 GGCCTGGAAGGCCCGGCCGTCACGGTGGACACCGCCTGCTCTTCCTCTCTGGTCGCCCTG 82204

Qy        541 CATCTGGCCTGTCAGGCGTTGCGTGCGGGTGAGTGCTCGATGGCGCTTGCCGGGGGTGTG 600
              ||   |||    |||||||||| ||||| |||||||||| ||||| | ||||| ||||||
Db      82205 CACTGGGCGGCGCAGGCGTTGCGGGCGGGGGAGTGCTCGTTGGCGTTGGCCGGTGGTGTG 82264

Qy        601 ACGGTGATGTCGTCTCCGGGTGCCTTCGTGGAGTTTTCGCGGCAGCGGGGTCTGGCCGCG 660
              |||||||||||| | ||| |  | |||||||||| ||||||||||||||||||||  |
Db      82265 ACGGTGATGTCGACGCCGAGCACGTTCGTGGAGTTCTCGCGGCAGCGGGGTCTGGCGCCT 82324

Qy        661 GACGGGCATTGCAAGGCGTTCTCGGCGGCGGCGGACGGGACCGGCTGGGGTGAGGGTGTG 720
              || || | ||  ||||||||| | | ||| ||||||||    | ||||   ||||| ||

Db      82325 GATGGTCGTTCGAAGGCGTTCGCCGAGGCCGCGGACGGCGTGGCCTGGTCCGAGGGCGTC 82384

Qy        721 GGGATGCTGCTGGTGGAGCGGCTCTCCGACGCCCATCGCAACGGTCACCGTGTCCTGGCC 780
              || |||||| |  |||||||||  || |||||    ||||||||||||     ||||||||
Db      82385 GGCATGCTGGTCCTGGAGCGGCAGTCGGACGCGGTGCGCAACGGTCACGAGATCCTGGCC 82444

Qy        781 GTGGTGCGTGGCAGTGCGGTCAACCAGGACGGTGCGAGCAACGGTCTGACCGCGCCCAAC 840
              |||||||| |||  ||||||||||||||||||||  ||||||||||||||||||||||||
Db      82445 GTGGTGCGCGGCTCGGCGGTCAACCAGGACGGTGCGTCCAACGGTCTGACCGCGCCCAAC 82504

Qy        841 GGGCCGTCCCAGCAGCGTGTCATCCGCCAGGCCCTCGCCAACGCCGGCTTGTCGGCCGGT 900
              || |||||||||||||| || ||||| |||||  | ||||  | ||||  ||||  | |
Db      82505 GGCCCGTCCCAGCAGCGGGTGATCCGTCAGGCGTTGGCCAGTGGCGGCCTGTCCACGGCC 82564

Qy        901 GATGTCGACGCGGTGGAGGCCCACGGCACCGGCACCACTTTGGGCGACCCGATCGAGGCC 960
              || || ||||| || ||||| |||||||| || || ||  | || ||||||||||||||||
Db      82565 GACGTGGACGCCGTTGAGGCGCACGGCACGGGTACGACGCTCGGTGACCCGATCGAGGCC 82624

Qy        961 CAGGCCCTCCTCGCGACCTACGGACAGGACCGTGCCGGCGAGGGGCCGCTGTGGCTGGGC 1020
              ||||| ||||| || |||||||| |  ||||| | |  ||||   ||||||  ||| |||
Db      82625 CAGGCGCTCCTGGCCACCTACGGTCGCGACCGCGACCCCGAGAACCCGCTGCTGCTCGGC 82684

Qy       1021 TCGGTCAAGTCCAATGTCGGTCACACACAGGCTGCCGCGGGCGTCGCCGGGGTGATCAAG 1080
              ||| ||||||||||  ||||||||| ||||| || || || |||||||| || ||||||
Db      82685 TCGATCAAGTCCAACATCGGTCACACCCAGGCAGCGGCCGGTGTCGCCGGTGTCATCAAG 82744

Qy       1081 ATGGTGATGGCGCTGCGGCATGGTCTGCTGCCGCGGACGTTGCATGTGGATGAGCCGTCG 1140
              ||||| |||||| ||||||| ||  ||||||||| |||  ||||||| || | ||||||
Db      82745 ATGGTCATGGCGATGCGGCACGGCGTGCTGCCGCAGACCCTGCATGTCGACGCGCCGTCC 82804

Qy       1141 CCGCATGTGGACTGGTCCGCGGGTGCGGTGCAGCTGCTGACGGAGACGGTGCCCTGGCCC 1200
              |||| || || |||  ||  || || ||  | ||||| || |||  |    |||||||
Db      82805 TCGCACGTCGATTGGAGCGTCGGCGCCGTCGAACTGCTCACCGAGCAGACCGCCTGGCC- 82863

Qy       1201 GGCGGGGAGGGGCGGCTACGGCGGGCAGGAGTGTCATCATTCGGCGTCAGCGGCACCAAC 1260
                | |   || |||    || || || || || || || || |||||| ||||||||||||||
Db      82864 --GGAGACCGGCCGGGCCCGTCGCGCCGGTGTCTCCTCCTTCGGCATCAGCGGCACCAAC 82921

Qy       1261 GCCCACGTCATCCTCGAAGAAGCACCCGCC 1290
              ||||||||| || ||||  |  | ||  ||
Db      82922 GCCCACGTCGTCATCGAGCAGTCCCCGACC 82951


RESULT 11
AX211739
LOCUS       AX211739               125401 bp    DNA     linear   PAT 06-SEP-2001
DEFINITION  Sequence 35 from Patent WO0159126.
ACCESSION   AX211739
VERSION     AX211739.1  GI:15523950
KEYWORDS    .
SOURCE      Streptomyces noursei.
  ORGANISM  Streptomyces noursei
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1  (bases 1 to 125401)
  AUTHORS   Zotchev,S.B., Sekurova,O.N., Fjaervik,E., Brautaset,T.,
            Stroem,A.R., Valla,S., Ellingsen,T.E., Sletta,H.V. and
            Gulliksen,O.M.
  TITLE     Gene cluster encoding a nystatin polyketide synthase and its
            manipulation and utility
  JOURNAL   Patent: WO 0159126-A 35 16-AUG-2001;
            Norges Teknisk Naturvitenskapelige Universitet (NO) ; STIFTELSEN
            IND OG TEKNISK FORSKNING VED NORGES TEKNISKE HOGSKOLE (NO) ;
            ALPHARMA AS (NO) ; Sinvent AS (NO) ; Zotchev, Sergey Borisovich
            (NO) ; Sekurova, Olga Nikolayivna (NO) ; Fjaervik, Espen (NO) ;
            Brautaset, Trygve (NO) ; Strom, Arne Reidar (NO) ; Valla, Svein
            (NO)
FEATURES             Location/Qualifiers
     source          1. .125401
                     /organism="Streptomyces noursei"
                     /db_xref="taxon:1971"
                     /note="ATCC 11455"
BASE COUNT    15664 a  49692 c  42871 g  17174 t
```

ORIGIN

Query Match 55.5%; Score 716; DB 6; Length 125401;
Best Local Similarity 73.3%; Pred. No. 4e-80;
Matches 946; Conservative 0; Mismatches 335; Indels 9; Gaps 2;

```
Qy          1 GAGCCGATCGCCATCGTTGGCATGGCCTGTCGTTTCCCCGGCGGAGTGACCTCGGCGGAC 60
              || ||||||| |||||| |||||||||||| || | |||||||||   |  ||||  | ||
Db      82079 GACCCGATCGTCATCGTCGGCATGGCCTGCCGCTACCCCGGCGGCATCGCCTCACCCGAG 82138

Qy         61 GACTTCTGGGATCTGATCTCCTCCGAGCAGGACGCGATCGGCGGATTCCCCACCGACCGC 120
              ||| |||||   ||| ||  |   |     |||||  |  |||   ||||||||| |||||
Db      82139 GACCTCTGGCGCCTGGTCAGCCAGGGCGCCGACGCCACTGGCCCGTTCCCCACCAACCGC 82198

Qy        121 GGCTGGGACCTGGACACGCTCTACGACCCCGACCCCGACCACCCCGGCACCTGCTACACC 180
              ||||||||||||||||  |||||||||||||||||||||||| ||| |||  |  | ||  |
Db      82199 GGCTGGGACCTGGACAACCTCTACGACCCCGACCCCGACCGCCCGGGCCGCACCCACGTC 82258

Qy        181 CGAAACGGCGGATTCCTCTACGACGCAGGCCACTTCGACGCCGAATTCTTCGGCATCAGC 240
              ||   |||||| |||||  ||||||| |||  |||||||||||| |||||||| || |||
Db      82259 CGCGCCGGCGGCTTCCTGCACGACGCCGGCTCCTTCGACGCCGACTTCTTCGGGATGAGC 82318

Qy        241 CCCCGCGAAGCCCTCGCCATGGACCCCCAGCAACGACTCCTCCTCGAAACCGCCTGGGAA 300
              || ||||| ||  | ||||  ||| |||||||  || || || ||||||   | ||||||||
Db      82319 CCGCGCGAGGCGATGGCCACCGACTCCCAGCAGCGCCTGCTGCTCGAACTCTCCTGGGAA 82378

Qy        301 ACCATCGAACACGCCGGCATCAACCCCCACACCCTCCACGGCACCCCCACCGGAGTCTTC 360
              || |||||| |||||||||| |||||  |  | ||  |  ||  |  ||   |||||| ||||||
Db      82379 GCCGTCGAACGCGCCGGCATCGACCCCGCCTCACTGCGCGACTCCGGCACCGGCGTCTTC 82438

Qy        361 ACCGGCACCAACGGACAGGACTACGCACTTCGCGTGCACAACGCGGGCCAGTCAACCGAT 420
              |||||  ||       | |||||||      ||      |   | ||| |       |||
Db      82439 GCCGGCGTCATGTACAACGACTACG------GCACCACCCTGACCGGCGACGAGTACGAG 82492

Qy        421 GGTTTCGCACTGACCGGAACCGCCGGCAGCGTCATCTCCGGTCGTATCTCGTACACGTTT 480
              |  |||        | ||| | ||||   ||||||  ||||||  ||   |||| |||||   |
Db      82493 GCGTTCCGCGGCAACGGCAGCGCCCCGAGCGTCGCCTCCGGCCGCGTCTCCTACACCCTC 82552

Qy        481 GGTTTTGAGGGTCCTGCGGTGTCGGTGGACACGGCTTGTTCCTCGTCGTTGGTGGCTTTG 540
              ||  |  || || || ||  || |||||||||| || || || ||  ||  |||| ||  ||
Db      82553 GGCCTGGAAGGCCCGGCCGTCACGGTGGACACCGCCTGCTCTTCCTCTCTGGTCGCCCTG 82612

Qy        541 CATCTGGCCTGTCAGGCGTTGCGTGCGGGTGAGTGCTCGATGGCGCTTGCCGGGGGTGTG 600
              ||   |||    |||||||||| ||||| |||||||| ||||| | ||||| ||||||
Db      82613 CACTGGGCGGCGCAGGCGTTGCGGGCGGGGGAGTGCTCGTTGGCGTTGGCCGGTGGTGTG 82672

Qy        601 ACGGTGATGTCGTCTCCGGGTGCCTTCGTGGAGTTTTCGCGGCAGCGGGGTCTGGCCGCG 660
              |||||||||||  | |||  |  | |||||||||||| ||||||||||||||||||||  |
Db      82673 ACGGTGATGTCGACGCCGAGCACGTTCGTGGAGTTCTCGCGGCAGCGGGGTCTGGCGCCT 82732

Qy        661 GACGGGCATTGCAAGGCGTTCTCGGCGGCGGCGGACGGGACCGGCTGGGGTGAGGGTGTG 720
              || || | ||  |||||||| | | ||| |||||||    | ||||   |||| ||
Db      82733 GATGGTCGTTCGAAGGCGTTCGCCGAGGCCGCGGACGGCGTGGCCTGGTCCGAGGGCGTC 82792

Qy        721 GGGATGCTGCTGGTGGAGCGGCTCTCCGACGCCCATCGCAACGGTCACCGTGTCCTGGCC 780
              || |||||| |  |||||||||  || |||||   |||||||||||    ||||||||
Db      82793 GGCATGCTGGTCCTGGAGCGGCAGTCGGACGCGGTGCGCAACGGTCACGAGATCCTGGCC 82852

Qy        781 GTGGTGCGTGGCAGTGCGGTCAACCAGGACGGTGCGAGCAACGGTCTGACCGCGCCCAAC 840
              |||||||| |||   ||||||||||||||||||||   ||||||||||||||||||||||
Db      82853 GTGGTGCGCGGCTCGGCGGTCAACCAGGACGGTGCGTCCAACGGTCTGACCGCGCCCAAC 82912

Qy        841 GGGCCGTCCCAGCAGCGTGTCATCCGCCAGGCCCTCGCCAACGCCGGCTTGTCGGCCGGT 900
              || |||||||||||| || |||||  |||||  | ||||  | ||||  ||||  | |
Db      82913 GGCCCGTCCCAGCAGCGGGTGATCCGTCAGGCGTTGGCCAGTGGCGGCCTGTCCACGGCC 82972

Qy        901 GATGTCGACGCGGTGGAGGCCCACGGCACCGGCACCACTTTGGGCGACCCGATCGAGGCC 960
              || ||||| || ||||| ||||||| || || ||  | || ||||||||||||||||||||
Db      82973 GACGTGGACGCCGTTGAGGCGCACGGCACGGGTACGACGCTCGGTGACCCGATCGAGGCC 83032

Qy        961 CAGGCCCTCCTCGCGACCTACGGACAGGACCGTGCCGGCGAGGGGCCGCTGTGGCTGGGC 1020
              ||||| ||||| || |||||||| |  |||| | |  |||   |||||  ||| |||
Db      83033 CAGGCGCTCCTGGCCACCTACGGTCGCGACCGCGACCCCGAGAACCCGCTGCTGCTCGGC 83092

Qy       1021 TCGGTCAAGTCCAATGTCGGTCACACACAGGCTGCCGCGGGCGTCGCCGGGGTGATCAAG 1080
              ||| ||||||||||  |||||||||| ||||| || || || |||||||| || ||||||
Db      83093 TCGATCAAGTCCAACATCGGTCACACCCAGGCAGCGGCCGGTGTCGCCGGTGTCATCAAG 83152

Qy       1081 ATGGTGATGGCGCTGCGGCATGGTCTGCTGCCGCGGACGTTGCATGTGGATGAGCCGTCG 1140
              ||||| |||||| ||||||| ||  |||||||||  |||  |||||||  ||  | ||||||
Db      83153 ATGGTCATGGCGATGCGGCACGGCGTGCTGCCGCAGACCCTGCATGTCGACGCGCCGTCC 83212

Qy       1141 CCGCATGTGGACTGGTCCGCGGGTGCGGTGCAGCTGCTGACGGAGACGGTGCCCTGGCCC 1200
              ||||  || || |||  ||  || || ||  | ||||| || |||  |    |||||||
Db      83213 TCGCACGTCGATTGGAGCGTCGGCGCCGTCGAACTGCTCACCGAGCAGACCGCCTGGCC- 83271

Qy       1201 GGCGGGGAGGGGCGGCTACGGCGGGCAGGAGTGTCATCATTCGGCGTCAGCGGCACCAAC 1260
                | |   || |||   || || || || || || || || |||||| ||||||||||||||
Db      83272 --GGAGACCGGCCGGGCCCGTCGCGCCGGTGTCTCCTCCTTCGGCATCAGCGGCACCAAC 83329

Qy       1261 GCCCACGTCATCCTCGAAGAAGCACCCGCC 1290
              ||||||||| || ||||  |  | ||  ||
Db      83330 GCCCACGTCGTCATCGAGCAGTCCCCGACC 83359
```

RESULT 12
AF079138
LOCUS AF079138 37948 bp DNA linear BCT 30-OCT-1998
DEFINITION Streptomyces venezuelae methymycin/pikromycin polyketide synthase gene cluster, complete sequence.
ACCESSION AF079138
VERSION AF079138.1 GI:3808326
KEYWORDS .
SOURCE Streptomyces venezuelae.
ORGANISM Streptomyces venezuelae
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae; Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE 1 (bases 1 to 37948)
AUTHORS Xue,Y., Zhao,L., Liu,H.W. and Sherman,D.H.
TITLE A gene cluster for macrolide antibiotic biosynthesis in Streptomyces venezuelae: architecture of metabolic diversity
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 95 (21), 12111-12116 (1998)
MEDLINE 98445333
PUBMED 9770448
REFERENCE 2 (bases 1 to 37948)
AUTHORS Xue,Y., Wilson,D., Zhuo,L. and Sherman,D.H.
TITLE Direct Submission
JOURNAL Submitted (17-JUL-1998) Department of Microbiology, University of Minnesota, 420 Delaware Street SE 1060, P.O. Box 196, Minneapolis, MN 55455, USA
REFERENCE 3 (bases 1 to 37948)
AUTHORS Xue,Y., Wilson,D., Zhuo,L. and Sherman,D.H.
TITLE Direct Submission
JOURNAL Submitted (29-OCT-1998) Department of Microbiology, University of Minnesota, 420 Delaware Street SE 1060, P.O. Box 196, Minneapolis, MN 55455, USA
REMARK Sequence update by submitter
COMMENT On Oct 30, 1998 this sequence version replaced gi:3800831.
FEATURES Location/Qualifiers
source 1. .37948
/organism="Streptomyces venezuelae"
/strain="ATCC15439"
/db_xref="ATCC:15439"
/db_xref="taxon:54571"
gene 154. .1122
/gene="pikR2"
CDS 154. .1122
/gene="pikR2"
/note="mls type resistance protein"
/codon_start=1
/transl_table=11
/product="rRNA methyltransferase PikR2"
/protein_id="AAC69327.1"
/db_xref="GI:3800832"
/translation="MAFSPQGGRHELGQNFLVDRSVIDEIDGLVARTKGPILEIGPGD
GALTLPLSRHGRPITAVELDGRRAQRLGARTPGHVTVVHHDFLQYPLPRNPHVVVGNV

PFHLTTAIMRRLLDAQHWHTAVLLVQWEVARRRAGVGGSTLLTAGWAPWYEFDLHSRV
PARAFRPMPGVDGGVLAIRRRSAPLVGQVKTYQDFVRQVFTGKGNGLKEILRRTGRIS
QRDLATWLRRNEISPHALPKDLKPGQWASLWELTGGTADGSFDGTAGGGAAGSHGAAR
VGAGHPGGRVSASRRGVPQARRGRGHAVRSSTGTEPRWGRGRAESA"

gene complement(1283. .2293)
/gene="pikR1"

CDS complement(1283. .2293)
/gene="pikR1"
/note="mls type resistance protein"
/codon_start=1
/transl_table=11
/product="rRNA methyltransferase PikR1"
/protein_id="AAC69328.1"
/db_xref="GI:3800833"
/translation="MAMRDSIPRRADRDTLRRELGQNFLQDDRAVRNLVTHVEGDGRN
VLEIGPGKGAITEELVRSFDTVTVVEMDPHWAAHVRRKFEGERVTVFQGDFLDFRIPR
DIDTVVGNVPFGITTQILRSLLESTNWQSAALIVQWEVARKRAGRSGGSLLTTSWAPW
YEFAVHDRVRASSFRPMPRVDGGVLTIRRRPQPLLPESASRAFQNFAEAVFTGPGRGL
AEILRRHIPKRTYRSLADRHGIPDGGLPKDLTLTQWIALFQASQPSYAPGAPGTRMPG
QGGGAGGRDYDSETSRAAVPGSRRYGPTRGGEPCAPRAQVRQTKGRQGARGSSYGRRT
GR"

gene 2912. .16753
/gene="pikAI"

CDS 2912. .16753
/gene="pikAI"
/note="multifunctional polyketide synthase; contains a
loading module, Pik module 1 and module 2"
/codon_start=1
/transl_table=11
/product="type I polyketide synthase PikAI"
/protein_id="AAC69329.1"
/db_xref="GI:3800834"
/translation="MSSAGITRTGARTPVTGRGAAAWDTGEVRVRRGLPPAGPDHAEH
SFSRAPTGDVRAELIRGEMSTVSKSESEEFVSVSNDAGSAHGTAEPVAVVGISCRVPG
ARDPREFWELLAAGGQAVTDVPADRWNAGDFYDPDRSAPGRSNSRWGGFIEDVDRFDA
AFFGISPREAAEMDPQQRLALELGWEALERAGIDPSSLTGTRTGVFAGAIWDDYATLK
HRQGGAAITPHTVTGLHRGIIANRLSYTLGLRGPSMVVDSGQSSSLVAVHLACESLRR
GESELALAGGVSLNLVPDSIIGASKFGGLSPDGRAYTFDARANGYVRGEGGGFVVLKR
LSRAVADGDPVLAVIRGSAVNNGGAAQGMTTPDAQAQEAVLREAHERAGTAPADVRYV
ELHGTGTPVGDPIEAAALGAALGTGRPAGQPLLVGSVKTNIGHLEGAAGIAGLIKAVL
AVRGRALPASLNYETPNPAIPFEELNLRVNTEYLPWEPEHDGQRMVVGVSSFGMGGTN
AHVVLEEAPGGCRGASVVESTVGGSAVGGGVVPWVVSAKSAAALDAQIERLAAFASRD
RTDGVDAGAVDAGAVDAGAVARVLAGGRAQFEHRAVVVGSGPDDLAAALAAPEGLVRG
VASGVGRVAFVFPGQGTQWAGMGAELLDSSAVFAAAMAECEAALSPYVDWSLEAVVRQ
APGAPTLERVDVVQPVTFAVMVSLARVWQHHGVTPQAVVGHSQGEIAAAYVAGALSLD
DAARVVTLRSKSIAAHLAGKGGMLSLALSEDAVLERLAGFDGLSVAAVNGPTATVVSG
DPVQIEELARACEADGVRARVIPVDYASHSRQVEIIESELAEVLAGLSPQAPRVPFFS
TLEGAWITEPVLDGGYWYRNLRHRVGFAPAVETLATDEGFTHFVEVSAHPVLTMALPG
TVTGLATLRRDNGGQDRLVASLAEAWANGLAVDWSPLLPSATGHHSDLPTYAFQTERH
WLGEIEALAPAGEPAVQPAVLRTEAAEPAELDRDEQLRVILDKVRAQTAQVLGYATGG
QIEVDRTFREAGCTSLTGVDLRNRINAAFGVRMAPSMIFDFPTPEALAEQLLLVVHGE
AAANPAGAEPAPVAAAGAVDEPVAIVGMACRLPGGVASPEDLWRLVAGGGDAISEFPQ
DRGWDVEGLYHPDPEHPGTSYVRQGGFIENVAGFDAAFFGISPREALAMDPQQRLLLE
TSWEAVEDAGIDPTSLRGRQVGVFTGAMTHEYGPSLRDGGEGLDGYLLTGNTASVMSG
RVSYTLGLEGPALTVDTACSSSLVALHLAVQALRKGEVDMALAGGVAVMPTPGMFVEF
SRQRGLAGDGRSKAFAASADGTSWSEGVGVLLVERLSDARRNGHQVLAVVRGSALNQD
GASNGLTAPNGPSQQRVIRRALADARLTTSDVDVVEAHGTGTRLGDPIEAQALIATYG
QGRDDEQPLRLGSLKSNIGHTQAAAGVSGVIKMVQAMRHGLLPKTLHVDEPSDQIDWS
AGAVELLTEAVDWPEKQDGGLRRAAVSSFGISGTNAHVVLEEAPVVVEGASVVEPSVG
GSAVGGGVTPWVVSAKSAAALDAQIERLAAFASRDRTDDADAGAVDAGAVAHVLADGR
AQFEHRAVALGAGADDLVQALADPDGLIRGTASGVGRVAFVFPGQGTQWAGMGAELLD
SSAVFAAAMAECEAALSPYVDWSLEAVVRQAPGAPTLERVDVVQPVTFAVMVSLARVW
QHHGVTPQAVVGHSQGEIAAAYVAGALPLDDAARVVTLRSKSIAAHLAGKGGMLSLAL
NEDAVLERLSDFDGLSVAAVNGPTATVVSGDPVQIEELAQACKADGFRARIIPVDYAS
HSRQVEIIESELAQVLAGLSPQAPRVPFFSTLEGTNAVVLRERVSDARRNGHQVLAVV
PAIETLAVDEGFTHFVEVSAHPVLTMTLPETVTGLGTLRREQGGQERLVTSLAEAWVN
GLPVAWTSLLPATASRPGLPTYAFQAERYWLENTPAALATGDDWRYRIDWKRLPAAEG
SERTGLSGRWLAVTPEDHSAQAAAVLTALVDAGAKVEVLTAGADDDREALAARLTALT
TGDGFTGVVSLLDGLVPQVAWVQALGDAGIKAPLWSVTQGAVSVGRLDTPADPDRAML
WGLGRVVALEHPERWAGLVDLPAQPDAAALAHLVTALSGATGEDQIAIRTTGLHARRL
ARAPLHGRRPTRDWQPHGTVLITGGTGALGSHAARWMAHHGAEHLLLVSRSGEQAPGA
TQLTAELTASGARVTIAACDVADPHAMRTLLDAIPAETPLTAVVHTAGALDDGIVDTL
TAEQVRRAHRAKAVGASVLDELTRDLDLDAFVLFSSVSSTLGIPGQGNYAPHNAYLDA

LAARRRATGRSAVSVAWGPWDGGGMAAGDGVAERLRNHGVPGMDPELALAALESALGR
DETAITVADIDWDRFYLAYSSGRPQPLVEELPEVRRIIDARDSATSGQGGSSAQGANP
LAERLAAAAPGERTEILLGLVRAQAAAVLRMRSPEDVAADRAFKDIGFDSLAGVELRN
RLTRATGLQLPATLVFDHPTPLALVSLLRSEFLGDEETADARRSAALPATVGAGAGAG
AGTDADDDPIAIVAMSCRYPGDIRSPEDLWRMLSEGGEGITPFPTDRGWDLDGLYDAD
PDALGRAYVREGGFLHDAAEFDAEFFGVSPREALAMDPQQRMLLTTSWEAFERAGIEP
ASLRGSSTGVFIGLSYQDYAARVPNAPRGVEGYLLTGSTPSVASGRIAYTFGLEGPAT
TVDTACSSSLTALHLAVRALRSGECTMALAGGVAMMATPHMFVEFSRQRALAPDGRSK
AFSADADGFGAAEGVGLLLVERLSDARRNGHPVLAVVRGTAVNQDGASNGLTAPNGPS
QQRVIRQALADARLAPGDIDAVETHGTGTSLGDPIEAQGLQATYGKERPAERPLAIGS
VKSNIGHTQAAAGAAGIIKMVLAMRHGTLPKTLHADEPSPHVDWANSGLALVTEPIDW
PAGTGPRRAAVSSFGISGTNAHVVLEQAPDAAGEVLGADEVPEVSETVAMAGTAGTSE
VAEGSEASEAPAAPGSREASLPGHLPWVLSAKDEQSLRGQAAALHAWLSEPAADLSDA
DGPARLRDVGYTLATSRTAFAHRAAVTAADRDGFLDGLATLAQGGTSAHVHLDTARDG
TTAFLFTGQGSQRPGAGRELYDRHPVFARALDEICAHLDGHLELPLLDVMFAAEGSAE
AALLDETRYTQCALFALEVALFRLVESWGMRPAALLGHSVGEIAAAHVAGVFSLADAA
RLVAARGRLMQELPAGGAMLAVQAAEDEIRVWLETEERYAGRLDVAAVNGPEAAVLSG
DADAAREAEAYWSGLGRRTRALRVSHAFHSAHMDGMLDGFRAVLETVEFRRPSLTVVS
NVTGLAAGPDDLCDPEYWVRHVRGTVRFLDGVRVLRDLGVRTCLELGPDGVLTAMAAD
GLADTPADSAAGSPVGSPAGSPADSAAGALRPRPLLVALLRRKRSETETVADALGRAH
AHGTGPDWHAWFAGSGAHRVDLPTYSFRRDRYWLDAPAADTAVDTAGLGLGTADHPLL
GAVVSLPDRDGLLLTGRLSLRTHPWLADHAVLGSVLLPGAAMVELAAHAAESAGLRDV
RELTLLEPLVLPEHGGVELRVTVGAPAGEPGGESAGDGARPVSLHSRLADAPAGTAWS
CHATGLLATDRPELPVAPDRAAMWPPQGAEEVPLDGLYERLDGNGLAFGPLFQGLNAV
WRYEGEVFADIALPATTNATAPATANGGGSAAAAPYGIHPALLDASLHAIAVGGLVDE
PELVRVPFHWSGVTVHAAGAAAARVRLASAGTDAVSLSLTDGEGRPLVSVERLTLRPV
TADQAAASRVGGLMHRVAWRPYALASSGEQDPHATSYGPTAVLGKDELKVAAALESAG
VEVGLYPDLAALSQDVAAGAPAPRTVLAPLPAGPADGGAEGVRGTVARTLELLQAWLA
DEHLAGTRLLLVTRGAVRDPEGSGADDGGEDLSHAAAWGLVRTAQTENPGRFGLLDLA
DDASSYRTLPSVLSDAGLRDEPQLALHDGTIRLARLASVRPETGTAAPALAPEGTVLL
TGGTGGLGGLVARHVVGEWGVRRLLLVSRRGTDAPGADELVHELEALGADVSVAACDV
ADREALTAVLDAIPAEHPLTAVVHTAGVLSDGTLPSMTTEDVEHVLRPKVDAAFLLDE
LTSTPAYDLAAFVMFSSAAAVFGGAGQGAYAAANATLDALAWRRRAAGLPALSLGWGL
WAETSGMTGELGQADLRRMSRAGIGGISDAEGIALLDAALRDDRHPVLLPLRLDAAGL
RDAAGNDPAGIPALFRDVVGARTVRARPSAASASTTAGTAGTPGTADGAAETAAVTLA
DRAATVDGPARQRLLLEFVVGEVAEVLGHARGHRIDAERGFLDLGFDSLTAVELRNRL
NSAGGLALPATLVFDHPSPAALASHLDAELPRGASDQDGAGNRNGNENGTTASRSTAE
TDALLAQLTRLEGALVLTGLSDAPGSEEVLEHLRSLRSMVTGETGTGTASGAPDGAGS
GAEDRPWAAGDGAGGGSEDGAGVPDFMNASAEELFGLLDQDPSTD"

gene 16858. .28077
/gene="pikAII"

CDS 16858. .28077
/gene="pikAII"
/note="multifunctional polyketide synthase; contains Pik
module 3 and module 4"
/codon_start=1
/transl_table=11
/product="type I polyketide synthase PikAII"
/protein_id="AAC69330.1"
/db_xref="GI:3800835"
/translation="MSTVNEEKYLDYLRRATADLHEARGRLRELEAKAGEPVAIVGMA
CRLPGGVASPEDLWRLVAGGEDAISEFPQDRGWDVEGLYDPNPEATGKSYAREAGFLY
EAGEFDADFFGISPREALAMDPQQRLLLEASWEAFEHAGIPAATARGTSVGVFTGVMY
HDYATRLTDVPEGIEGYLGTGNSGSVASGRVAYTLGLEGPAVTVDTACSSSLVALHLA
VQALRKGEVDMALAGGVTVMSTPSTFVEFSRQRGLAPDGRSKSFSSTADGTSWSEGVG
VLLVERLSDARRKGHRILAVVRGTAVNQDGASSGLTAPNGPSQQRVIRRALADARLTT
SDVDVVEAHGTGTRLGDPIEAQAVIATYGQGRDGEQPLRLGSLKSNIGHTQAAAGVSG
VIKMVQAMRHGVLPKTLHVEKPTDQVDWSAGAVELLTEAMDWPDKGDGGLRRAAVSSF
GVSGTNAHVVLEEAPAAEETPASEATPAVEPSVGAGLVPWLVSAKTPAALDAQIGRLA
AFASQGRTDAADPGAVARVLAGGRAEFEHRAVVLGTGQDDFAQALTAPEGLIRGTPSD
VGRVAFVFPGQGTQWAGMGAELLDVSKEFAAAMAECESALSRYVDWSLEAVVRQAPGA
PTLERVDVVQPVTFAVMVSLAKVWQHHGVTPQAVVGHSQGEIAAAYVAGALTLDDAAR
VVTLRSKSIAAHLAGKGGMISLALSEEATRQRIENLHGLSIAAVNGPTATVVSGDPTQ
IQELAQACEADGVRARIIPVDYASHSAHVETIESELAEVLAGLSPRTPEVPFFSTLEG
AWITEPVLDGTYWYRNLRHRVGFAPAVETLATDEGFTHFIEVSAHPVLTMTLPETVTG
LGTLRREQGGQERLVTSLAEAWTNGLTIDWAPVLPTATGHHPELPTYAFQRRHYWLHD
SPAVQGSVQDSWRYRIDWKRLAVADASERAGLSGRWLVVVPEDRSAEAAPVLAALSGA
GADPVQLDVSPLGDRQRLAATLGEALAAAGGAVDGVLSLLAWDESAHPGHPAPFTRGT
GATLTLVQALEDAGVAAPLWCVTHGAVSVGRADHVTSPAQAMVWGMGRVAALEHPERW
GGLIDLPSDADRAALDRMTTVLAGGTGEDQVAVRASGLLARRLVRASLPAHGTASPWW
QADGTVLVTGAEEPAAAEAARRLARDGAGHLLLHTTPSGSEGAEGTSGAAEDSGLAGL
VAELADLGATATVVTCDLTDAEAAARLLAGVSDAHPLSAVLHLPPTVDSEPLAATDAD
ALARVVTAKATAALHLDRLLRAAAAGGRPPVLVLFSSVAAIWGGAGQGAYAAGTAFL

```
                 DALAGQHRADGPTVTSVAWSPWEGSRVTEGATGERLRRLGLRPLAPATALTALDTALG
                 HGDTAVTIADVDWSSFAPGFTTARPGTLLADLPEARRALDEQQSTTAADDTVLSRELG
                 ALTGAEQQRRMQELVREHLAVVLNHPSPEAVDTGRAFRDLGFDSLTAVELRNRLKNAT
                 GLALPATLVFDYPTPRTLAEFLLAEILGEQAGAGEQLPVDGGVDDEPVAIVGMACRLP
                 GGVASPEDLWRLVAGGEDAISGFPQDRGWDVEGLYDPDPDASGRTYCRAGGFLDEAGE

 Query Match          55.5%;  Score 715.4;  DB 1;  Length 37948;
 Best Local Similarity  72.6%;  Pred. No. 6.3e-80;
 Matches  939;  Conservative    0;  Mismatches  351;  Indels    3;  Gaps    1;

Qy          1 GAGCCGATCGCCATCGTTGGCATGGCCTGTCGTTTCCCCGGCGGAGTGACCTCGGCGGAC 60
              ||||| |||| ||||| |||||||| || ||  | || ||||| ||  ||||| ||||
Db      21475 GAGCCCGTCGCGATCGTCGGCATGGCGTGCCGCCTGCCGGGCGGTGTCGCCTCGCCGGAG 21534

Qy         61 GACTTCTGGGATCTGATCTCCTCCGAGCAGGACGCGATCGGCGGATTCCCCACCGACCGC 120
              ||| | |||   ||| |  ||  ||   |||||||||||  ||| |||||     ||||||
Db      21535 GACCTGTGGCGGCTGGTGGCCGGCGGCGAGGACGCGATCTCCGGCTTCCCGCAGGACCGC 21594

Qy        121 GGCTGGGACCTGGACACGCTCTACGACCCCGACCCCGACCACCCCGGCACCTGCTACACC 180
              |||||||| ||||   ||| |||||||| ||||| |||    |||         |||  |
Db      21595 GGCTGGGACGTGGAGGGGCTGTACGACCCGGACCCGGACGCGTCCGGGCGGACGTACTGC 21654

Qy        181 CGAAACGGCGGATTCCTCTACGACGCAGGCCACTTCGACGCCGAATTCTTCGGCATCAGC 240
              ||   ||| || |||||| |||| || ||| | ||||||||||| ||||||||  |||
Db      21655 CGTGCCGGTGGCTTCCTCGACGAGGCGGGCGAGTTCGACGCCGACTTCTTCGGGATCTCG 21714

Qy        241 CCCCGCGAAGCCCTCGCCATGGACCCCCAGCAACGACTCCTCCTCGAAACCGCCTGGGAA 300
              || ||||| |||||||||||||||||| ||||| || |||||||| || ||| |||||||
Db      21715 CCGCGCGAGGCCCTCGCCATGGACCCGCAGCAGCGGCTCCTCCTGGAGACCTCCTGGGAG 21774

Qy        301 ACCATCGAACACGCCGGCATCAACCCCCACACCCTCCACGGCACCCCCACCGGAGTCTTC 360
              || ||||  |||||| ||| ||||   | |||| || ||    |   ||| || |||
Db      21775 GCCGTCGAGGACGCCGGGATCGACCCGACCTCCCTTCAGGGGCAGCAGGTCGGCGTGTTC 21834

Qy        361 ACCGGCACCAACGGACAGGACTACGCACTTCGCGTGCACAACGCGGGCCAGTCAACCGAT 420
              | |||||||||||  |   ||||||  |   |  | |||| | | | ||      ||
Db      21835 GCGGGCACCAACGGCCCCCACTACGAGCCGCTGCTCCGCAACACCGCCGAGGATCTTGAG 21894

Qy        421 GGTTTCGCACTGACCGGAACCGCCGGCAGCGTCATCTCCGGTCGTATCTCGTACACGTTT 480
              |||| ||    ||| || | ||||| |||| |||| || || ||| |||||||||  |
Db      21895 GGTTACGTCGGGACGGGCAACGCCGCCAGCATCATGTCGGGCCGTGTCTCGTACACCCTC 21954

Qy        481 GGTTTTGAGGGTCCTGCGGTGTCGGTGGACACGGCTTGTTCCTCGTCGTTGGTGGCTTTG 540
              ||  | ||||| || || ||  |||| ||||| || || ||||| ||| |||| ||  ||
Db      21955 GGCCTGGAGGGCCCGGCCGTCACGGTCGACACCGCCTGCTCCTCCTCGCTGGTCGCCCTG 22014

Qy        541 CATCTGGCCTGTCAGGCGTTGCGTGCGGGTGAGTGCTCGATGGCGCTTGCCGGGGGTGTG 600
              || || |||   |||||  ||||   ||| || |||    ||||||| || || ||||||
Db      22015 CACCTCGCCGTGCAGGCCCTGCGCAAGGGCGAATGCGGACTGGCGCTCGCGGGCGGTGTG 22074

Qy        601 ACGGTGATGTCGTCTCCGGGTGCCTTCGTGGAGTTTTCGCGGCAGCGGGGTCTGGCCGCG 660
              ||||| |||||| | ||     | ||||||||||    |||||||| || || || | |
Db      22075 ACGGTCATGTCGACGCCCACGACGTTCGTGGAGTTCAGCCGGCAGCGCGGGCTCGCGGAG 22134

Qy        661 GACGGGCATTGCAAGGCGTTCTCGGCGGCGGCGGACGGGACCGGCTGGGGTGAGGGTGTG 720
              ||||| |  | |||||||||| | ||| ||||||||||    ||||  ||  |||||  ||
Db      22135 GACGGCCGGTCGAAGGCGTTCGCCGCGTCGGCGGACGGCTTCGGCCCGGCGGAGGGCGTC 22194

Qy        721 GGGATGCTGCTGGTGGAGCGGCTCTCCGACGCCCATCGCAACGGTCACCGTGTCCTGGCC 780
              || ||||| || || ||||| || || |||||||  |||||||| |||||||| |||||
Db      22195 GGCATGCTCCTCGTCGAGCGCCTGTCGGACGCCCGCCGCAACGGACACCGTGTGCTGGCG 22254

Qy        781 GTGGTGCGTGGCAGTGCGGTCAACCAGGACGGTGCGAGCAACGGTCTGACCGCGCCCAAC 840
              || ||||| ||||| |||||||||||||||| |||||||||| |||||||| || |||
Db      22255 GTCGTGCGCGGCAGCGCGGTCAACCAGGACGGCGCGAGCAACGGCCTGACCGCCCCGAAC 22314

Qy        841 GGGCCGTCCCAGCAGCGTGTCATCCGCCAGGCCCTCGCCAACGCCGGCTTGTCGGCCGGT 900
              ||||| || ||||||| |||||||| |  || |||||  |||||  |  || || |||
Db      22315 GGGCCCTCGCAGCAGCGCGTCATCCGGCGCGCGCTCGCGGACGCCCGACTGACGACCGCC 22374

Qy        901 GATGTCGACGCGGTGGAGGCCCACGGCACCGGCACCACTTTGGGCGACCCGATCGAGGCC 960
              || || ||||  || |||||||||||||| |||||     | |||||||||||||||||
Db      22375 GACGTGGACGTCGTCGAGGCCCACGGCACGGGCACGCGACTCGGCGACCCGATCGAGGCA 22434

Qy        961 CAGGCCCTCCTCGCGACCTACGGACAGGACCGTGCCGGCGAGGGGCCGCTGTGGCTGGGC 1020
              ||||||||| |||| |||||||| ||||  || | |  |||    ||||||| | |||||
Db      22435 CAGGCCCTCATCGCCACCTACGGCCAGGGGCGCGACACCGAACAGCCGCTGCGCCTGGGG 22494

Qy       1021 TCGGTCAAGTCCAATGTCGGTCACACACAGGCTGCCGCGGGCGTCGCCGGGGTGATCAAG 1080
              ||| | ||||||||  |||| |||||  ||||| |||| || ||| ||||  | ||||||
Db      22495 TCGTTGAAGTCCAACATCGGACACACCCAGGCCGCCGCCGGTGTCTCCGGCATCATCAAG 22554

Qy       1081 ATGGTGATGGCGCTGCGGCATGGTCTGCTGCCGCGGACGTTGCATGTGGATGAGCCGTCG 1140
              |||||   |||| |||| || ||  | ||||||  |||| | || |||||   |||||||
Db      22555 ATGGTCCAGGCGATGCGCCACGGCGTCCTGCCGAAGACGCTCCACGTGGACCGGCCGTCG 22614

Qy       1141 CCGCATGTGGACTGGTCCGCGGGTGCGGTGCAGCTGCTGACGGAGACGGTGCCCTGGCCC 1200
                ||  | |||||||| |||||  ||||  |||||||  || ||| |  ||  ||||||
Db      22615 GACCAGATCGACTGGTCGGCGGGCACGGTCGAGCTGCTCACCGAGGCCATGGACTGGCCG 22674

Qy       1201 GG---CGGGGAGGGGCGGCTACGGCGGGCAGGAGTGTCATCATTCGGCGTCAGCGGCACC 1257
              ||      |||||  |||| || || || |  || || || |||||| ||||||||||
Db      22675 AGGAAGCAGGAGGGCGGGCTGCGCCGCGCGGCCGTCTCCTCCTTCGGCATCAGCGGCACG 22734

Qy       1258 AACGCCCACGTCATCCTCGAAGAAGCACCCGCC 1290
              ||||| ||| || | |||||||||||| || | |
Db      22735 AACGCGCACATCGTGCTCGAAGAAGCCCCGGTC 22767
```

```
RESULT 13
AX089460
LOCUS       AX089460               9513 bp    DNA     linear   PAT 21-MAR-2001
DEFINITION  Sequence 45 from Patent WO0116303.
ACCESSION   AX089460
VERSION     AX089460.1  GI:13443721
KEYWORDS    .
SOURCE      Saccharopolyspora spinosa.
  ORGANISM  Saccharopolyspora spinosa
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Pseudonocardineae; Pseudonocardiaceae;
            Saccharopolyspora.
REFERENCE   1  (bases 1 to 9513)
  AUTHORS   Eberz,G., Moehrle,V., Froede,R., Velten,R. and Salas,J.A.
  TITLE     Nucleic acids which code for the enzyme activities of the spinosyn
            biosynthesis
  JOURNAL   Patent: WO 0116303-A 45 08-MAR-2001;
            BAYER AG (DE)
FEATURES             Location/Qualifiers
     source          1. .9513
                     /organism="Saccharopolyspora spinosa"
                     /db_xref="taxon:60894"
     CDS             1. .9513
                     /note="ORF20; Polyketidsynthase"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAC34893.1"
                     /db_xref="GI:13443722"
                     /translation="MSNEEKLREYLRRALVDLHQARERLHEAESGEREPIAIVAMGCR
                     YPGGVQDPEGLWKLVASGGDAIGEFPADRGWHLDELYDPDPDQPGTCYTRHGGFLHDA
                     GEFDAGFFDISPREALAMDPQQRLLLEISWETVESAGMDPRSLRGSRTGVFAGLMYEG
                     YDTGAHRAGEGVEGYLGTGNAGSVASGRVAYAFGFEGPAVTVDTACSSSLVALHLACQ
                     SLRQGECDLALAGGVTVMSTPERFVEFSRQRGLAPDGRCKSFAAAADGTGWGEGAGLV
                     LLERLSDARRNGHRVLAVVRGSAVNQDGASNGLTAPNGLAQERVIQQVLTSAGLSASD
                     VDAVEAHGTGTRLGDPIEAQALIAAYGQDRDRDRPLWLGSVKSNIGHTQAAAGVAGVI
                     KMVMAMRHGELPRTLHVDEPNSHVDWSAGAVRLLTENIRWPGTGTRRAGVSSFGVSGT
                     NAHVILEHDPLAVTENEEAAQSPAPGIVPWALSGRSSTALRAQAERLRELCEQTDPDP
                     VDVGFSLAATRTAWEHRAVVLGRDSATLRSGLGVVASGEPAVDVVEGSVLDGEVVFVF
                     PGQGWQWAGMAVDLLDASPTFARHMDECATALRRYVDWSLVDVLRGAENSPPLDRVDV
                     LQPASFAVMVSLAEVWRSYGVRPAAVVGHSQGEIAAACAAGVLPLEDAARLVALRSRA
                     LKGLSGRGGMASLACPADEVAALFAGSGGRLEVAAINGPRSVVVSGDLEAVDELLAEC
                     AEKDMRARRIPVDYASHSAHVEVVRSPVLAAAAGVRHRDGQVPWWSTVIGDWVDPARL
                     DGEYWYRNLRQPVRFEHAVQGLVERGFGLFIEMSAHPVLTTAVEETGAESETAVAAVG
                     TLRRDSGGLRRLLHSLAEAYVRGATVDWAVAFGGAGRRLDLPTYPFQRQRYWLDKGAA
                     SDEARAVSDPAAGWFWQAVARQDLKSVSDALDLDADAPLSATLPALSVWHRQERERVL
                     ADGWRYRVDWVRVAPQPVRRTRETWLLVVPPGGIEEALVERLTDALNTRGISTLRLDV
                     PPAATSGELATELRAAADGDPVKAILSLTALDERPHPECKDVPSGIALLLNLVKALGE
```

```
                     ADLRIPLWTITRGAVKAGPADRLLRPMQAQAWGLGRVAALEHPERWGGLIDLPDSLDG
                     DVLTRLGEALTNGLAEDQLAIRQSGVLARRLVPAPANQPAGRKWRPRGSALITGGLGA
                     VGAQVARWLAEIGAERIVLTSRRGNQAAGAAELEAELRALGAQVSIVACDVTDRAEMS
                     ALLAEFDVTAVFHAAGVGRLLPLAETDQNGLAEICAAKVRGAQVLDELCDSTDLDAFV
                     LFSSGAGVWGGGGQGAYGAANAFLDTLAEQRRARGLPATSISWGSWAGGGMADGAAGE
                     HLRRRGIRPMPAASAILALQEVLDQDETCVSIADVDWDRFVPTFAATRATRLFDEVPA
                     ARKAMPANGPAEPGGSPFARNLAELPEAQRRHELVDLVCAQVATVLGHGSREEVQPER
                     AFRALGFDSLMAVDLRNRLTTATGLRLPTTTVFDYPNPAALAAHLLEELVGDVASAAV
                     TAASAPASDEPIAIVAMSCRFPGGAHSPEDLWRLVAAGTEVIGEFPSDRGWDAEGLYD
                     PDASRPGTTYARMAGFLYDAGEFDADLFGISPREALAMDPQQRLVLEIAWEALERAGI
                     DPLSLKGSGVGTYIGAGSRGYATDVRQFPEEAEGYLLTGTSASVLSGRVAYSFGFEGP
                     AVTVDTACSSSLVALHLACQSLRSGECDLALAGGVTVMSTPEMFVEFSRQRGLAPDGR
                     CKSFAESADGTGWGEGAGLLLLERLSDAHRNGHRVLAVVRGSAVNQDGASNGLAAPNG
                     PSQQRVINQALANAALSASDVDAVEAHGTGTRLGDPIEAQALIATYGQARERDRPLWL
                     GSVKSNIGHTQAAAGVAGVIKMVMAMRHGQLPASLHADEPTSEVDWSSGAVRLLAEQV
                     PWPESDRVRRVGVSSFGISGTNAHVILEQATNAPDSTAETDKTESGSTVDIPVVPWLV
                     SGKTTDSLRGQAERVLSQVESRPEQRSLDVAYSLASGRAALDERAVVLGADRGELVAG
                     LAALAAGQEASGVISGTRASARFGFVFSGQGGQWLGMGRALYSKFPVFAAAFDEACAE
                     LEAHLGEDRRVRDVVFGSDAQLLDQTLWAQSGLFALQAGLLGLLGSWGVRPDVVMGHS
                     VGELAAAFAAGVLSLRDAARLVAARARLMQALPSDGAMLAVAAGEDLVRPLLAGREES
                     VSVAALNAPGSVVLSGDREVLASIVGRLTELRVRTRRLRVSHAFHSHRMDPMLGEFAQ
                     IAESAEFGKPTTPLVSTLTGELDRAAEMSTPGYWVRQAREPVRFADGVQALAAQGIGT
                     VVELGPDGTLAALVRECATESDRVGRISSIPLMRRERDETRSVMTALAHLHTRGGEVD
                     WQAFFAGTGARQLELPTYAFQRQHYWIESSARPARDRADIGEVAEQFWTAVDQGDLAT
                     LVAALDLGADDDTCASLSDVLPALSSWRSGLRNRSLVDSCRYRISWHSSREVPAPKIS
                     GTWLLVVPGAADDGLVTALTSSLVGGGAEVVRIGLSEEDPHREDVAQRLANALTDAGQ
                     LGGVLSLLGLDESPAPGFSCLPTGFALTVQLLRALRKADVEAPFWAVTRGGVALEDVR
                     VSPEQALVWGLLRVAGLEHPEFWGGLIDLPSDWDDRLGARLAGVLADGGEDQVAIRRG
                     GVFVRRLERAGASGAGSVWRPRGTVLVTGGTGGLGAHVARWLAGAGAEHVVLTSRRGA
                     DAPGAGELRAELEALGARVSIVPCDVADRDAVAGVLAGIGGECPLTAVVHAAGVGEAG
                     DVVEMGLADFAAVLSAKVRGAANLDELLADSELDAFVMFSSVSGVWGAGGQGAYAAAN
                     AYLDALAEQRRARGLVGTAVAWGPWAGDGMAAGETGAQLHRMGLASMEPSAALLALQG
                     ALDRDETSLVVADVDWARFAPAFTSARRRPLLDTIDEARAALETTGEQAGTGKPVELT
                     QRLAGLSRKERDDAVLDLVRAETAAVLGRDDATALAPSRPFQELGFDSLMAVELRNRL
                     NTATGIQLPASTIFDYPNAESLSRHLCAELFPTETTVDSALAELDRIEQQLSMLTGEA
                     RARDRIATRLRALHEKWNSAAEVPTGADVLSTLDSATHDEIFEFIDNELDLS"
BASE COUNT     1392 a   2759 c   3637 g   1725 t
ORIGIN

  Query Match           55.4%;  Score 715.2;  DB 6;  Length 9513;
  Best Local Similarity   72.9%;  Pred. No. 9.5e-80;
  Matches  937;  Conservative    0;  Mismatches  343;  Indels    6;  Gaps    1;

Qy          1 GAGCCGATCGCCATCGTTGGCATGGCCTGTCGTTTCCCCGGCGGAGTGACCTCGGCGGAC 60
              || || ||||| ||||| |  |||| ||| || | ||| || || |||      ||||
Db        100 GAACCCATCGCGATCGTGGCGATGGGCTGCCGGTACCCGGGTGGGGTGCAGGACCCGGAA 159

Qy         61 GACTTCTGGGATCTGATCTCCTCCGAGCAGGACGCGATCGGCGGATTCCCCACCGACCGC 120
              |   | ||| | ||| || ||||||     ||||| ||||| | ||||||| | || ||
Db        160 GGGCTGTGGAAACTGGTCGCCTCCGGTGGCGACGCCATCGGTGAATTCCCCGCTGATCGT 219

Qy        121 GGCTGGGACCTGGACACGCTCTACGACCCCGACCCCGACCACCCCGGCACCTGCTACACC 180
              || ||| |||| ||| |||||||| |||||||| || || ||||| ||||||||||||||
Db        220 GGTTGGCACCTCGACGAGCTCTACGATCCCGACCCGGATCAGCCCGGAACCTGCTACACC 279

Qy        181 CGAAACGGCGGATTCCTCTACGACGCAGGCCACTTCGACGCCGAATTCTTCGGCATCAGC 240
              ||  ||||||| |||||| ||||||| ||| | |||||||| | |||||||| |||||||
Db        280 CGGCACGGCGGCTTCCTCCACGACGCCGGCGAGTTCGACGCGGGATTCTTCGACATCAGC 339

Qy        241 CCCCGCGAAGCCCTCGCCATGGACCCCCAGCAACGACTCCTCCTCGAAACCGCCTGGGAA 300
              ||||| || || ||||| |||||||| ||||| || || || || |||| | |||||||
Db        340 CCCCGTGAGGCGCTCGCGATGGACCCGCAGCAGCGGCTGCTGCTGGAAATCTCCTGGGAG 399

Qy        301 ACCATCGAACACGCCGGCATCAACCCCCACACCCTCCACGGCACCCCCACCGGAGTCTTC 360
              ||| |||||  ||| || ||  ||||     || | |  || | || |||||| || |||
Db        400 ACCGTCGAATCCGCTGGGATGGACCCGAGGTCCTTGCGGGGGAGCCGCACCGGGGTGTTC 459

Qy        361 ACCGGCACCAACGGACAGGACTACGCACTTCGCGTGCACAACGCGGGCCAGTCAACCGAT 420
                | ||    |     ||| ||| |     |||  |||    || ||   |      |||
Db        460 GCGGGATTGATGTACGAGGGCTATGACACCGGCGCCCACCGGGCAGGAGAAGGTGTCGAA 519

Qy        421 GGTTTCGCACTGACCGGAACCGCCGGCAGCGTCATCTCCGGTCGTATCTCGTACACGTTT 480
              || |        ||||| |  || || ||||||  ||| |||||  |  ||||  ||||

Db        520 GGCTATCTCGGAACCGGCAATGCGGGAAGCGTCGCCTCTGGTCGGGTTGCGTATGCGTTC 579

Qy        481 GGTTTTGAGGGTCCTGCGGTGTCGGTGGACACGGCTTGTTCCTCGTCGTTGGTGGCTTTG 540
              || || ||||| || |||||| |||| |||||||| || || ||||||||||||||  ||
Db        580 GGGTTCGAGGGCCCAGCGGTGACGGTAGACACGGCGTGCTCGTCGTCGTTGGTGGCGCTG 639

Qy        541 CATCTGGCCTGTCAGGCGTTGCGTGCGGGTGAGTGCTCGATGGCGCTTGCCGGGGGTGTG 600
              ||| |||| |||||| |||||||     ||| |||||      ||||||| ||||| || |||
Db        640 CATTTGGCGTGTCAGTCGTTGCGGCAGGGCGAGTGTGATCTGGCGCTGGCCGGTGGAGTG 699

Qy        601 ACGGTGATGTCGTCTCCGGGTGCCTTCGTGGAGTTTTCGCGGCAGCGGGGTCTGGCCGCG 660
              ||||||||||| | ||||     |||||||||| || || ||||| ||||| ||  ||
Db        700 ACGGTGATGTCGACGCCGGAGAGGTTCGTGGAGTTCTCCCGTCAGCGTGGTCTCGCACCG 759

Qy        661 GACGGGCATTGCAAGGCGTTCTCGGCGGCGGCGGACGGGACCGGCTGGGGTGAGGGTGTG 720
              || |||  || ||| ||||| ||||||| ||||| || ||||| ||||||||||||||
Db        760 GATGGGCGGTGTAAGTCGTTCGCGGCGGCTGCGGATGGAACCGGTTGGGGTGAGGGTGCC 819

Qy        721 GGGATGCTGCTGGTGGAGCGGCTCTCCGACGCCCATCGCAACGGTCACCGTGTCCTGGCC 780
              || || || || |||||||||| || ||||||   || ||||| || || || |||||
Db        820 GGTTTGGTGTTGCTGGAGCGGCTGTCAGACGCCAGGCGGAACGGGCATCGGGTACTGGCG 879

Qy        781 GTGGTGCGTGGCAGTGCGGTCAACCAGGACGGTGCGAGCAACGGTCTGACCGCGCCCAAC 840
              || || ||||| || ||||| || ||||||||||||   |||||  |||| || || |||
Db        880 GTTGTTCGTGGTAGCGCGGTGAATCAGGACGGTGCGTCGAACGGATTGACGGCCCCGAAC 939

Qy        841 GGGCCGTCCCAGCAGCGTGTCATCCGCCAGGCCCTCGCCAACGCCGGCTTGTCGGCCGGT 900
              |||| | ||||| |||| ||||| |  ||||   ||| | |  || ||  ||||||
Db        940 GGGCTGGCCCAGGAGCGGGTCATTCAGCAGGTGCTCACGAGTGCGGGGCTGTCGGCGTCC 999

Qy        901 GATGTCGACGCGGTGGAGGCCCACGGCACCGGCACCACTTTGGGCGACCCGATCGAGGCC 960
              ||||| |||| |||||||| || || || || ||      | || || ||||||||||
Db       1000 GATGTGGACGCTGTGGAGGCGCATGGAACGGGTACGCGGCTTGGTGATCCGATCGAGGCG 1059

Qy        961 CAGGCCCTCCTCGCGACCTACGGACAGGACCGTGCCGGCGAGGGGCCGCTGTGGCTGGGC 1020
              ||||| ||  | ||  |||| |||||||| || | | | ||  |||||||||| ||||
Db       1060 CAGGCTCTGATAGCCGCCTATGGACAGGATCGGGACCGGGACCGGCCGCTGTGGTTGGGG 1119

Qy       1021 TCGGTCAAGTCCAATGTCGGTCACACACAGGCTGCCGCGGGCGTCGCCGGGGTGATCAAG 1080
              |||||||||||||  ||||||| || |||| || |||||||||| || |||||||||
Db       1120 TCGGTCAAGTCCAACATCGGTCATACGCAGGCGGCTGCGGGCGTCGCTGGTGTGATCAAG 1179

Qy       1081 ATGGTGATGGCGCTGCGGCATGGTCTGCTGCCGCGGACGTTGCATGTGGATGAGCCGTCG 1140
              ||||| |||||| ||||||| ||   |||||||| |||||||| |||| ||||||
Db       1180 ATGGTCATGGCGATGCGGCACGGGGAGCTGCCGCGCACGTTGCACGTGGACGAGCCGAAT 1239

Qy       1141 CCGCATGTGGACTGGTCCGCGGGTGCGGTGCAGCTGCTGACGGAGACGGTGCCCTGGCCC 1200
              ||||  |||||||||| || |||||||  |  || |||||  |||    | | ||||||
Db       1240 TCGCACGTGGACTGGTCGGCTGGTGCGGTCCGACTCCTGACCGAGAACATCCGCTGGCCA 1299

Qy       1201 GGCGGGGAGGGGCGGCTACGGCGGGCAGGAGTGTCATCATTCGGCGTCAGCGGCACCAAC 1260
              ||      | | |    || || || |||||||| || |||||  || |||||  ||||||
Db       1300 GG------GACGGGTACGCGCCGCGCTGGAGTGTCGTCGTTCGGGGTAAGCGGTACCAAC 1353

Qy       1261 GCCCACGTCATCCTCGAAGAAGCACC 1286
              || ||||||||||||||| | |  ||
Db       1354 GCACACGTCATCCTCGAACACGACCC 1379


RESULT 14
AX089417
LOCUS       AX089417               50000 bp    DNA     linear   PAT 21-MAR-2001
DEFINITION  Sequence 2 from Patent WO0116303.
ACCESSION   AX089417
VERSION     AX089417.1  GI:13443678
KEYWORDS    .
SOURCE      Saccharopolyspora spinosa.
  ORGANISM  Saccharopolyspora spinosa
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Pseudonocardineae; Pseudonocardiaceae;
            Saccharopolyspora.
REFERENCE   1  (bases 1 to 50000)
  AUTHORS   Eberz,G., Moehrle,V., Froede,R., Velten,R. and Salas,J.A.
```

```
  TITLE     Nucleic acids which code for the enzyme activities of the spinosyn
            biosynthesis
  JOURNAL   Patent: WO 0116303-A 2 08-MAR-2001;
            BAYER AG (DE)
FEATURES             Location/Qualifiers
     source          1. .50000
                     /organism="Saccharopolyspora spinosa"
                     /db_xref="taxon:60894"
BASE COUNT     6867 a  14165 c  19274 g   9694 t
ORIGIN

  Query Match           55.4%;  Score 715.2;  DB 6;  Length 50000;
  Best Local Similarity  72.9%;  Pred. No. 6.3e-80;
  Matches  937;  Conservative    0;  Mismatches  343;  Indels    6;  Gaps    1;

Qy          1 GAGCCGATCGCCATCGTTGGCATGGCCTGTCGTTTCCCCGGCGGAGTGACCTCGGCGGAC 60
              || || ||||| ||||| |  |||| ||| || | ||| || || |||       ||||
Db      10147 GAACCCATCGCGATCGTGGCGATGGGCTGCCGGTACCCGGGTGGGGTGCAGGACCCGGAA 10206

Qy         61 GACTTCTGGGATCTGATCTCCTCCGAGCAGGACGCGATCGGCGGATTCCCCACCGACCGC 120
              |   | ||| | ||| || ||||||     ||||| ||||| | ||||||| | || ||
Db      10207 GGGCTGTGGAAACTGGTCGCCTCCGGTGGCGACGCCATCGGTGAATTCCCCGCTGATCGT 10266

Qy        121 GGCTGGGACCTGGACACGCTCTACGACCCCGACCCCGACCACCCCGGCACCTGCTACACC 180
              || ||| |||| ||| |||||||| |||||||| || || ||||| ||||||||||||
Db      10267 GGTTGGCACCTCGACGAGCTCTACGATCCCGACCCGGATCAGCCCGGAACCTGCTACACC 10326

Qy        181 CGAAACGGCGGATTCCTCTACGACGCAGGCCACTTCGACGCCGAATTCTTCGGCATCAGC 240
              ||  ||||||| |||||| ||||||| ||| | |||||||| | |||||||| |||||||
Db      10327 CGGCACGGCGGCTTCCTCCACGACGCCGGCGAGTTCGACGCGGGATTCTTCGACATCAGC 10386

Qy        241 CCCCGCGAAGCCCTCGCCATGGACCCCCAGCAACGACTCCTCCTCGAAACCGCCTGGGAA 300
              ||||| || || ||||| |||||||| ||||| || || || || |||| | |||||||
Db      10387 CCCCGTGAGGCGCTCGCGATGGACCCGCAGCAGCGGCTGCTGCTGGAAATCTCCTGGGAG 10446

Qy        301 ACCATCGAACACGCCGGCATCAACCCCCACACCCTCCACGGCACCCCCACCGGAGTCTTC 360
              ||| ||||| ||| || ||  ||||     || | |  || | || |||||| || |||
Db      10447 ACCGTCGAATCCGCTGGGATGGACCCGAGGTCCTTGCGGGGGAGCCGCACCGGGGTGTTC 10506

Qy        361 ACCGGCACCAACGGACAGGACTACGCACTTCGCGTGCACAACGCGGGCCAGTCAACCGAT 420
              | ||   |     ||| ||| |     |||  |||   || ||  |   ·  ||||
Db      10507 GCGGGATTGATGTACGAGGGCTATGACACCGGCGCCCACCGGGCAGGAGAAGGTGTCGAA 10566

Qy        421 GGTTTCGCACTGACCGGAACCGCCGGCAGCGTCATCTCCGGTCGTATCTCGTACACGTTT 480
              || |         ||||| |  || || ||||||  ||| |||||  |  ||||  ||||
Db      10567 GGCTATCTCGGAACCGGCAATGCGGGAAGCGTCGCCTCTGGTCGGGTTGCGTATGCGTTC 10626

Qy        481 GGTTTTGAGGGTCCTGCGGTGTCGGTGGACACGGCTTGTTCCTCGTCGTTGGTGGCTTTG 540
              || || ||||| || |||||| |||| |||||||| || || ||||||||||||||  ||
Db      10627 GGGTTCGAGGGCCCAGCGGTGACGGTAGACACGGCGTGCTCGTCGTCGTTGGTGGCGCTG 10686

Qy        541 CATCTGGCCTGTCAGGCGTTGCGTGCGGGTGAGTGCTCGATGGCGCTTGCCGGGGGTGTG 600
              ||| |||| |||||| |||||||   ||| |||||     ||||||| ||||| || |||
Db      10687 CATTTGGCGTGTCAGTCGTTGCGGCAGGGCGAGTGTGATCTGGCGCTGGCCGGTGGAGTG 10746

Qy        601 ACGGTGATGTCGTCTCCGGGTGCCTTCGTGGAGTTTTCGCGGCAGCGGGGTCTGGCCGCG 660
              |||||||||||| | ||||      |||||||||||| || || ||||| ||||| ||  ||
Db      10747 ACGGTGATGTCGACGCCGGAGAGGTTCGTGGAGTTCTCCCGTCAGCGTGGTCTCGCACCG 10806

Qy        661 GACGGGCATTGCAAGGCGTTCTCGGCGGCGGCGGACGGGACCGGCTGGGGTGAGGGTGTG 720
              || ||||  || ||| ||||| |||||| ||||| || ||||| ||||||||||||||
Db      10807 GATGGGCGGTGTAAGTCGTTCGCGGCGGCTGCGGATGGAACCGGTTGGGGTGAGGGTGCC 10866

Qy        721 GGGATGCTGCTGGTGGAGCGGCTCTCCGACGCCCATCGCAACGGTCACCGTGTCCTGGCC 780
              ||  || || || |||||||||| || ||||||   || ||||| || || || |||||
Db      10867 GGTTTGGTGTTGCTGGAGCGGCTGTCAGACGCCAGGCGGAACGGGCATCGGGTACTGGCG 10926

Qy        781 GTGGTGCGTGGCAGTGCGGTCAACCAGGACGGTGCGAGCAACGGTCTGACCGCGCCCAAC 840
              || || ||||| || ||||| || ||||||||||||    |||||  |||| || || |||
Db      10927 GTTGTTCGTGGTAGCGCGGTGAATCAGGACGGTGCGTCGAACGGATTGACGGCCCCGAAC 10986

Qy        841 GGGCCGTCCCAGCAGCGTGTCATCCGCCAGGCCCTCGCCAACGCCGGCTTGTCGGCCGGT 900
              |||| | ||||| |||| ||||| |  ||||  ||| | |  || ||  |||||||
Db      10987 GGGCTGGCCCAGGAGCGGGTCATTCAGCAGGTGCTCACGAGTGCGGGGCTGTCGGCGTCC 11046

Qy        901 GATGTCGACGCGGTGGAGGCCCACGGCACCGGCACCACTTTGGGCGACCCGATCGAGGCC 960
              ||||| ||||| |||||||| || || || || || |      | || || ||||||||||
Db      11047 GATGTGGACGCTGTGGAGGCGCATGGAACGGGTACGCGGCTTGGTGATCCGATCGAGGCG 11106

Qy        961 CAGGCCCTCCTCGCGACCTACGGACAGGACCGTGCCGGCGAGGGGCCGCTGTGGCTGGGC 1020
              ||||| ||  | ||  |||| |||||||| || | | | ||  |||||||||| ||||
Db      11107 CAGGCTCTGATAGCCGCCTATGGACAGGATCGGGACCGGGACCGGCCGCTGTGGTTGGGG 11166

Qy       1021 TCGGTCAAGTCCAATGTCGGTCACACACAGGCTGCCGCGGGCGTCGCCGGGGTGATCAAG 1080
              |||||||||||||  ||||||| || ||||| || |||||||||| || |||||||||
Db      11167 TCGGTCAAGTCCAACATCGGTCATACGCAGGCGGCTGCGGGCGTCGCTGGTGTGATCAAG 11226

Qy       1081 ATGGTGATGGCGCTGCGGCATGGTCTGCTGCCGCGGACGTTGCATGTGGATGAGCCGTCG 1140
              ||||| |||||| ||||||| ||   |||||||||| |||||||| ||||| ||||||
Db      11227 ATGGTCATGGCGATGCGGCACGGGGAGCTGCCGCGCACGTTGCACGTGGACGAGCCGAAT 11286

Qy       1141 CCGCATGTGGACTGGTCCGCGGGTGCGGTGCAGCTGCTGACGGAGACGGTGCCCTGGCCC 1200
              ||||  |||||||||||  || |||||||| |  ||  ||||| ||||   | | ||||||
Db      11287 TCGCACGTGGACTGGTCGGCTGGTGCGGTCCGACTCCTGACCGAGAACATCCGCTGGCCA 11346

Qy       1201 GGCGGGGAGGGGCGGCTACGGCGGGCAGGAGTGTCATCATTCGGCGTCAGCGGCACCAAC 1260
              ||       |  |  |   || || || |||||||| || ||||| || ||||| ||||||
Db      11347 GG------GACGGGTACGCGCCGCGCTGGAGTGTCGTCGTTCGGGGTAAGCGGTACCAAC 11400

Qy       1261 GCCCACGTCATCCTCGAAGAAGCACC 1286
              || ||||||||||||||  |  |  ||
Db      11401 GCACACGTCATCCTCGAACACGACCC 11426


RESULT 15
AX089420
LOCUS       AX089420               50000 bp    DNA     linear   PAT 21-MAR-2001
DEFINITION  Sequence 5 from Patent WO0116303.
ACCESSION   AX089420
VERSION     AX089420.1  GI:13443681
KEYWORDS    .
SOURCE      Saccharopolyspora spinosa.
  ORGANISM  Saccharopolyspora spinosa
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Pseudonocardineae; Pseudonocardiaceae;
            Saccharopolyspora.
REFERENCE   1  (bases 1 to 50000)
  AUTHORS   Eberz,G., Moehrle,V., Froede,R., Velten,R. and Salas,J.A.
  TITLE     Nucleic acids which code for the enzyme activities of the spinosyn
            biosynthesis
  JOURNAL   Patent: WO 0116303-A 5 08-MAR-2001;
            BAYER AG (DE)
FEATURES             Location/Qualifiers
     source          1. .50000
                     /organism="Saccharopolyspora spinosa"
                     /db_xref="taxon:60894"
BASE COUNT     6944 a  14312 c  19170 g   9574 t
ORIGIN

  Query Match           55.4%;  Score 715.2;  DB 6;  Length 50000;
  Best Local Similarity  72.9%;  Pred. No. 6.3e-80;
  Matches  937;  Conservative    0;  Mismatches  343;  Indels    6;  Gaps    1;

Qy          1 GAGCCGATCGCCATCGTTGGCATGGCCTGTCGTTTCCCCGGCGGAGTGACCTCGGCGGAC 60
              || || ||||| ||||| |  |||| ||| || | ||| || || |||       ||||
Db      14523 GAACCCATCGCGATCGTGGCGATGGGCTGCCGGTACCCGGGTGGGGTGCAGGACCCGGAA 14582

Qy         61 GACTTCTGGGATCTGATCTCCTCCGAGCAGGACGCGATCGGCGGATTCCCCACCGACCGC 120
              |   | ||| | ||| || ||||||     ||||| ||||| | ||||||| | || ||
Db      14583 GGGCTGTGGAAACTGGTCGCCTCCGGTGGCGACGCCATCGGTGAATTCCCCGCTGATCGT 14642

Qy        121 GGCTGGGACCTGGACACGCTCTACGACCCCGACCCCGACCACCCCGGCACCTGCTACACC 180
              || ||| |||| |||  |||||||| |||||||| || || ||||| ||||||||||||
Db      14643 GGTTGGCACCTCGACGAGCTCTACGATCCCGACCCGGATCAGCCCGGAACCTGCTACACC 14702

Qy        181 CGAAACGGCGGATTCCTCTACGACGCAGGCCACTTCGACGCCGAATTCTTCGGCATCAGC 240
```

```
Db      14703 CGGCACGGCGGCTTCCTCCACGACGCCGGCGAGTTCGACGCGGGATTCTTCGACATCAGC 14762

Qy        241 CCCCGCGAAGCCCTCGCCATGGACCCCCAGCAACGACTCCTCCTCGAAACCGCCTGGGAA 300

Db      14763 CCCCGTGAGGCGCTCGCGATGGACCCGCAGCAGCGGCTGCTGCTGGAAATCTCCTGGGAG 14822

Qy        301 ACCATCGAACACGCCGGCATCAACCCCCACACCCTCCACGGCACCCCCACCGGAGTCTTC 360

Db      14823 ACCGTCGAATCCGCTGGGATGGACCCGAGGTCCTTGCGGGGGAGCCGCACCGGGGTGTTC 14882

Qy        361 ACCGGCACCAACGGACAGGACTACGCACTTCGCGTGCACAACGCGGGCCAGTCAACCGAT 420

Db      14883 GCGGGATTGATGTACGAGGGCTATGACACCGGCGCCCACCGGGCAGGAGAAGGTGTCGAA 14942

Qy        421 GGTTTCGCACTGACCGGAACCGCCGGCAGCGTCATCTCCGGTCGTATCTCGTACACGTTT 480

Db      14943 GGCTATCTCGGAACCGGCAATGCGGGAAGCGTCGCCTCTGGTCGGGTTGCGTATGCGTTC 15002

Qy        481 GGTTTTGAGGGTCCTGCGGTGTCGGTGGACACGGCTTGTTCCTCGTCGTTGGTGGCTTTG 540

Db      15003 GGGTTCGAGGGCCCAGCGGTGACGGTAGACACGGCGTGCTCGTCGTCGTTGGTGGCGCTG 15062

Qy        541 CATCTGGCCTGTCAGGCGTTGCGTGCGGGTGAGTGCTCGATGGCGCTTGCCGGGGGTGTG 600

Db      15063 CATTTGGCGTGTCAGTCGTTGCGGCAGGGCGAGTGTGATCTGGCGCTGGCCGGTGGAGTG 15122

Qy        601 ACGGTGATGTCGTCTCCGGGTGCCTTCGTGGAGTTTTCGCGGCAGCGGGGTCTGGCCGCG 660

Db      15123 ACGGTGATGTCGACGCCGGAGAGGTTCGTGGAGTTCTCCCGTCAGCGTGGTCTCGCACCG 15182

Qy        661 GACGGGCATTGCAAGGCGTTCTCGGCGGCGGCGGACGGGACCGGCTGGGGTGAGGGTGTG 720

Db      15183 GATGGGCGGTGTAAGTCGTTCGCGGCGGCTGCGGATGGAACCGGTTGGGGTGAGGGTGCC 15242

Qy        721 GGGATGCTGCTGGTGGAGCGGCTCTCCGACGCCCATCGCAACGGTCACCGTGTCCTGGCC 780

Db      15243 GGTTTGGTGTTGCTGGAGCGGCTGTCAGACGCCAGGCGGAACGGGCATCGGGTACTGGCG 15302

Qy        781 GTGGTGCGTGGCAGTGCGGTCAACCAGGACGGTGCGAGCAACGGTCTGACCGCGCCCAAC 840

Db      15303 GTTGTTCGTGGTAGCGCGGTGAATCAGGACGGTGCGTCGAACGGATTGACGGCCCCGAAC 15362

Qy        841 GGGCCGTCCCAGCAGCGTGTCATCCGCCAGGCCCTCGCCAACGCCGGCTTGTCGGCCGGT 900

Db      15363 GGGCTGGCCCAGGAGCGGGTCATTCAGCAGGTGCTCACGAGTGCGGGGCTGTCGGCGTCC 15422

Qy        901 GATGTCGACGCGGTGGAGGCCCACGGCACCGGCACCACTTTGGGCGACCCGATCGAGGCC 960

Db      15423 GATGTGGACGCTGTGGAGGCGCATGGAACGGGTACGCGGCTTGGTGATCCGATCGAGGCG 15482

Qy        961 CAGGCCCTCCTCGCGACCTACGGACAGGACCGTGCCGGCGAGGGGCCGCTGTGGCTGGGC 1020

Db      15483 CAGGCTCTGATAGCCGCCTATGGACAGGATCGGGACCGGGACCGGCCGCTGTGGTTGGGG 15542

Qy       1021 TCGGTCAAGTCCAATGTCGGTCACACACAGGCTGCCGCGGGCGTCGCCGGGGTGATCAAG 1080

Db      15543 TCGGTCAAGTCCAACATCGGTCATACGCAGGCGGCTGCGGGCGTCGCTGGTGTGATCAAG 15602

Qy       1081 ATGGTGATGGCGCTGCGGCATGGTCTGCTGCCGCGGACGTTGCATGTGGATGAGCCGTCG 1140

Db      15603 ATGGTCATGGCGATGCGGCACGGGGAGCTGCCGCGCACGTTGCACGTGGACGAGCCGAAT 15662

Qy       1141 CCGCATGTGGACTGGTCCGCGGGTGCGGTGCAGCTGCTGACGGAGACGGTGCCCTGGCCC 1200

Db      15663 TCGCACGTGGACTGGTCGGCTGGTGCGGTCCGACTCCTGACCGAGAACATCCGCTGGCCA 15722

Qy       1201 GGCGGGGAGGGGCGGCTACGGCGGGCAGGAGTGTCATCATTCGGCGTCAGCGGCACCAAC 1260

Db      15723 GG------GACGGGTACGCGCCGCGCTGGAGTGTCGTCGTTCGGGGTAAGCGGTACCAAC 15776

Qy       1261 GCCCACGTCATCCTCGAAGAAGCACC 1286
```

```
Db      15777 GCACACGTCATCCTCGAACACGACCC 15802
```

Search completed: June 17, 2003, 17:55:41
Job time : 2507.79 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 17, 2003, 11:26:45 ; Search time 256.616 Seconds
(without alignments)
11320.736 Million cell updates/sec

Title: US-09-914-286-1_COPY_6256_7545
Perfect score: 1290
Sequence: 1 gagccgatcgccatcgttgg.....tcctcgaagaagcacccgcc 1290

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_101002:*
1: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1290	100.0	11916	22	AAH79279	Streptomyces averm
2	1290	100.0	12381	21	AAZ58381	Streptomyces averm
3	1290	100.0	30690	21	AAA92301	S. avermitilis ave
4	1290	100.0	30690	22	AAH79277	Streptomyces averm
5	1032.2	80.0	31422	21	AAA92302	S. avermitilis ave
6	1032.2	80.0	31422	22	AAH79278	Streptomyces averm
7	732.2	56.8	44377	18	AAT78508	Platenolide syntha
8	732.2	56.8	44377	18	AAT80414	Platenolide syntha
9	722.6	56.0	13987	18	AAT80415	Hybrid srmG/tylG O
10	716	55.5	65140	22	AAD17184	Streptomyces nours
11	716	55.5	125401	22	AAD17186	Streptomyces nours
12	715.4	55.5	11220	21	AAZ87298	S. venezuelae macr
13	715.4	55.5	36778	21	AAZ87318	S. venezuelae pik
14	715.4	55.5	37948	21	AAZ87285	S. venezuelae pik
15	715.4	55.5	38506	21	AAA75633	Nucleotide sequenc
16	715.4	55.5	38506	21	AAZ56001	Recombinant cosmid
17	715.2	55.4	9513	22	AAF88337	S. spinosa DNA fra
18	715.2	55.4	50000	22	AAF88313	S. spinosa DNA fra
19	715.2	55.4	50000	22	AAF88316	S. spinosa DNA fra
20	715.2	55.4	80161	20	AAZ21501	DNA fragment of Sa
21	706	54.7	3978	21	AAC55785	Type I polyketide
22	706	54.7	12249	21	AAC55840	Complete Mitomycin
23	706	54.7	18331	21	AAC55857	Complete nucleotid
24	696	54.0	20394	22	AAF24892	Pimaricin biosynth
25	688.4	53.4	43280	18	AAT80413	Tylactone synthase
26	687.4	53.3	13842	21	AAZ87297	S. venezuelae macr
27	684.6	53.1	29879	14	AAQ46806	eryA region of S.
28	683.6	53.0	14775	22	AAF88338	S. spinosa DNA fra
29	680.4	52.7	15872	21	AAZ87283	S. venezuelae vep
30	678	52.6	27541	22	AAD17185	Streptomyces nours
31	675.6	52.4	15872	18	AAT68715	Streptomyces venez
32	675.2	52.3	5544	22	AAF24891	Pimaricin biosynth
33	674.2	52.3	53789	19	AAV21187	Amycolatopsis medi
34	673	52.2	4041	21	AAZ87300	S. venezuelae macr
35	673	52.2	50937	21	AAA09469	Streptococcus olea
36	668	51.8	4478	21	AAA14661	Nucleotide sequenc
37	668	51.8	4571	21	AAA14662	Nucleotide sequenc
38	664.4	51.5	77536	21	AAA14651	Nucleotide sequenc
39	662.8	51.4	4466	21	AAA14663	Nucleotide sequenc
40	662.8	51.4	4547	21	AAA14664	Nucleotide sequenc
41	656.8	50.9	5676	19	AAV21186	Amycolatopsis medi
42	655.2	50.8	4689	21	AAZ87299	S. venezuelae macr
43	653.6	50.7	47981	22	AAF30757	Micromonospora meg
c 44	653.2	50.6	77536	21	AAA14651	Nucleotide sequenc
45	650	50.4	16767	22	AAF88339	S. spinosa DNA fra

ALIGNMENTS

RESULT 1
AAH79279
ID AAH79279 standard; DNA; 11916 BP.
XX
AC AAH79279;
XX
DT 04-DEC-2001 (first entry)
XX
DE Streptomyces avermitilis coding sequence derivative SEQ ID NO: 3.
XX
KW Avermectin aglycone synthase; AAS; avermectin derivative;
KW drug production; vetinary drug; pesticide; ds.
XX
OS Synthetic.
XX
FH Key Location/Qualifiers
FT CDS 1..11916
FT /*tag= a
FT /product= "AAG65268"
FT /partial
XX
PN WO200162939-A1.
XX
PD 30-AUG-2001.
XX
PF 23-FEB-2001; 2001WO-JP01381.
XX
PR 24-FEB-2000; 2000JP-0047405.
XX
PA (KYOW) KYOWA HAKKO KOGYO KK.
PA (KITA) KITASATO INST.
XX

PI Endo H, Yamaguchi H, Kanda Y, Hashimoto S, Omura S, Ikeda H;
XX
DR WPI; 2001-582053/65.
DR P-PSDB; AAG65268.
XX
PT New modified avermectin aglycone synthase derived from Streptomyces
PT avermectilis used in production of 22,23-dihydroavermectin B1a used in
PT drugs and pesticides -
XX
PS Claim 10; Page 149-167; 257pp; Japanese.
XX
CC The present invention relates to the production of modified derivatives
CC of avermectin aglycone synthase (AAS) derived from Streptomyces
CC avermitilis. The activity of an acyl carrier protein (ACP),
CC beta-ketoacyl carrier protein synthase (KS), acyltransferase (AT),
CC beta-ketoacyl carrier protein reductase (KR), dehydratase (DH), enoyl
CC reductase (ER) and/or thioesterase (TE) domain may be reduced or
CC suppressed. The process can be used in the production of drugs, vetinary
CC drugs and pesticides. The present sequence is modified version of a
CC fragment of the S. avermitilis genome.
XX
SQ Sequence 11916 BP; 1831 A; 4428 C; 3820 G; 1837 T; 0 other;

Query Match 100.0%; Score 1290; DB 22; Length 11916;
Best Local Similarity 100.0%; Pred. No. 1.2e-206;
Matches 1290; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 GAGCCGATCGCCATCGTTGGCATGGCCTGTCGTTTCCCCGGCGGAGTGACCTCGGCGGAC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6256 GAGCCGATCGCCATCGTTGGCATGGCCTGTCGTTTCCCCGGCGGAGTGACCTCGGCGGAC 6315

Qy         61 GACTTCTGGGATCTGATCTCCTCCGAGCAGGACGCGATCGGCGGATTCCCCACCGACCGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6316 GACTTCTGGGATCTGATCTCCTCCGAGCAGGACGCGATCGGCGGATTCCCCACCGACCGC 6375

Qy        121 GGCTGGGACCTGGACACGCTCTACGACCCCGACCCCGACCACCCCGGCACCTGCTACACC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6376 GGCTGGGACCTGGACACGCTCTACGACCCCGACCCCGACCACCCCGGCACCTGCTACACC 6435

Qy        181 CGAAACGGCGGATTCCTCTACGACGCAGGCCACTTCGACGCCGAATTCTTCGGCATCAGC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6436 CGAAACGGCGGATTCCTCTACGACGCAGGCCACTTCGACGCCGAATTCTTCGGCATCAGC 6495

Qy        241 CCCCGCGAAGCCCTCGCCATGGACCCCCAGCAACGACTCCTCCTCGAAACCGCCTGGGAA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6496 CCCCGCGAAGCCCTCGCCATGGACCCCCAGCAACGACTCCTCCTCGAAACCGCCTGGGAA 6555

Qy        301 ACCATCGAACACGCCGGCATCAACCCCCACACCCTCCACGGCACCCCCACCGGAGTCTTC 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6556 ACCATCGAACACGCCGGCATCAACCCCCACACCCTCCACGGCACCCCCACCGGAGTCTTC 6615

Qy        361 ACCGGCACCAACGGACAGGACTACGCACTTCGCGTGCACAACGCGGGCCAGTCAACCGAT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6616 ACCGGCACCAACGGACAGGACTACGCACTTCGCGTGCACAACGCGGGCCAGTCAACCGAT 6675

Qy        421 GGTTTCGCACTGACCGGAACCGCCGGCAGCGTCATCTCCGGTCGTATCTCGTACACGTTT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6676 GGTTTCGCACTGACCGGAACCGCCGGCAGCGTCATCTCCGGTCGTATCTCGTACACGTTT 6735

Qy        481 GGTTTTGAGGGTCCTGCGGTGTCGGTGGACACGGCTTGTTCCTCGTCGTTGGTGGCTTTG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6736 GGTTTTGAGGGTCCTGCGGTGTCGGTGGACACGGCTTGTTCCTCGTCGTTGGTGGCTTTG 6795

Qy        541 CATCTGGCCTGTCAGGCGTTGCGTGCGGGTGAGTGCTCGATGGCGCTTGCCGGGGGTGTG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6796 CATCTGGCCTGTCAGGCGTTGCGTGCGGGTGAGTGCTCGATGGCGCTTGCCGGGGGTGTG 6855

Qy        601 ACGGTGATGTCGTCTCCGGGTGCCTTCGTGGAGTTTTCGCGGCAGCGGGGTCTGGCCGCG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6856 ACGGTGATGTCGTCTCCGGGTGCCTTCGTGGAGTTTTCGCGGCAGCGGGGTCTGGCCGCG 6915

Qy        661 GACGGGCATTGCAAGGCGTTCTCGGCGGCGGCGGACGGGACCGGCTGGGGTGAGGGTGTG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6916 GACGGGCATTGCAAGGCGTTCTCGGCGGCGGCGGACGGGACCGGCTGGGGTGAGGGTGTG 6975

Qy        721 GGGATGCTGCTGGTGGAGCGGCTCTCCGACGCCCATCGCAACGGTCACCGTGTCCTGGCC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6976 GGGATGCTGCTGGTGGAGCGGCTCTCCGACGCCCATCGCAACGGTCACCGTGTCCTGGCC 7035

Qy        781 GTGGTGCGTGGCAGTGCGGTCAACCAGGACGGTGCGAGCAACGGTCTGACCGCGCCCAAC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7036 GTGGTGCGTGGCAGTGCGGTCAACCAGGACGGTGCGAGCAACGGTCTGACCGCGCCCAAC 7095

Qy        841 GGGCCGTCCCAGCAGCGTGTCATCCGCCAGGCCCTCGCCAACGCCGGCTTGTCGGCCGGT 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7096 GGGCCGTCCCAGCAGCGTGTCATCCGCCAGGCCCTCGCCAACGCCGGCTTGTCGGCCGGT 7155

Qy        901 GATGTCGACGCGGTGGAGGCCCACGGCACCGGCACCACTTTGGGCGACCCGATCGAGGCC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7156 GATGTCGACGCGGTGGAGGCCCACGGCACCGGCACCACTTTGGGCGACCCGATCGAGGCC 7215

Qy        961 CAGGCCCTCCTCGCGACCTACGGACAGGACCGTGCCGGCGAGGGGCCGCTGTGGCTGGGC 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7216 CAGGCCCTCCTCGCGACCTACGGACAGGACCGTGCCGGCGAGGGGCCGCTGTGGCTGGGC 7275

Qy       1021 TCGGTCAAGTCCAATGTCGGTCACACACAGGCTGCCGCGGGCGTCGCCGGGGTGATCAAG 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7276 TCGGTCAAGTCCAATGTCGGTCACACACAGGCTGCCGCGGGCGTCGCCGGGGTGATCAAG 7335

Qy       1081 ATGGTGATGGCGCTGCGGCATGGTCTGCTGCCGCGGACGTTGCATGTGGATGAGCCGTCG 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7336 ATGGTGATGGCGCTGCGGCATGGTCTGCTGCCGCGGACGTTGCATGTGGATGAGCCGTCG 7395

Qy       1141 CCGCATGTGGACTGGTCCGCGGGTGCGGTGCAGCTGCTGACGGAGACGGTGCCCTGGCCC 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7396 CCGCATGTGGACTGGTCCGCGGGTGCGGTGCAGCTGCTGACGGAGACGGTGCCCTGGCCC 7455

Qy       1201 GGCGGGGAGGGGCGGCTACGGCGGGCAGGAGTGTCATCATTCGGCGTCAGCGGCACCAAC 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7456 GGCGGGGAGGGGCGGCTACGGCGGGCAGGAGTGTCATCATTCGGCGTCAGCGGCACCAAC 7515

Qy       1261 GCCCACGTCATCCTCGAAGAAGCACCCGCC 1290
              ||||||||||||||||||||||||||||||
Db       7516 GCCCACGTCATCCTCGAAGAAGCACCCGCC 7545
```

RESULT 2
AAZ58381
ID AAZ58381 standard; DNA; 12381 BP.
XX
AC AAZ58381;
XX
DT 23-MAY-2000 (first entry)
XX
DE Streptomyces avermitilis avermectin polyketide synthase modules 1+2.
XX
KW Polyketide synthase; avermectin; insecticide; ss.
XX
OS Streptomyces avermitilis.
XX
PN WO200001827-A2.
XX
PD 13-JAN-2000.
XX
PF 06-JUL-1999; 99WO-GB02158.
XX
PR 06-JUL-1998; 98GB-0014622.
XX
PA (BIOT-) BIOTICA TECHNOLOGY LTD.
PA (PFIZ) PFIZER INC.
XX
PI Kellenberger JL, Leadlay PF, Staunton J, Stutzman-Engwall KJ;
PI McArthur HAI;
XX
DR WPI; 2000-182117/16.
XX

```
PT   Mutated Type I polyketide synthase containing a polylinker site in an
PT   extension module for replacement of a reductive loop sequence, for
PT   producing polyketides, e.g. B1 avermectin -
XX
PS   Disclosure; Fig 7a-f; 75pp; English.
XX
CC   The present sequence is that of DNA encoding the first 2 modules
CC   of the avermectin polyketide synthase (PKS) of Streptomyces
CC   avermitilis.  The invention relates to nucleic acids encoding a
CC   Type I PKS such as avermectin in which a polylinker with multiple
CC   restriction sites replaces or 1 more PKS genes encoding enzymes
CC   associated with reduction.  Novel PKS are provided in which in
CC   which the reductive loop in a selected module of the Type I PKS is
CC   replaced with the equivalent segment from the same or different
CC   PKS gene cluster or by a mutated or synthetic segment.  Vectors and
CC   host cells, and methods for producing novel polyketides by
CC   culturing host cells are claimed.  The polyketides obtained are
CC   useful as antibiotics and insecticides.  Fermentation products
CC   containing C22-C23 dihydroavermectin, ivermectin and B1
CC   avermectins are claimed.
XX
SQ   Sequence 12381 BP; 1884 A; 4561 C; 4005 G; 1931 T; 0 other;

  Query Match           100.0%;  Score 1290;  DB 21;  Length 12381;
  Best Local Similarity  100.0%;  Pred. No. 1.2e-206;
  Matches 1290;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GAGCCGATCGCCATCGTTGGCATGGCCTGTCGTTTCCCCGGCGGAGTGACCTCGGCGGAC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7073 GAGCCGATCGCCATCGTTGGCATGGCCTGTCGTTTCCCCGGCGGAGTGACCTCGGCGGAC 7132

Qy         61 GACTTCTGGGATCTGATCTCCTCCGAGCAGGACGCGATCGGCGGATTCCCCACCGACCGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7133 GACTTCTGGGATCTGATCTCCTCCGAGCAGGACGCGATCGGCGGATTCCCCACCGACCGC 7192

Qy        121 GGCTGGGACCTGGACACGCTCTACGACCCCGACCCCGACCACCCCGGCACCTGCTACACC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7193 GGCTGGGACCTGGACACGCTCTACGACCCCGACCCCGACCACCCCGGCACCTGCTACACC 7252

Qy        181 CGAAACGGCGGATTCCTCTACGACGCAGGCCACTTCGACGCCGAATTCTTCGGCATCAGC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7253 CGAAACGGCGGATTCCTCTACGACGCAGGCCACTTCGACGCCGAATTCTTCGGCATCAGC 7312

Qy        241 CCCCGCGAAGCCCTCGCCATGGACCCCCAGCAACGACTCCTCCTCGAAACCGCCTGGGAA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7313 CCCCGCGAAGCCCTCGCCATGGACCCCCAGCAACGACTCCTCCTCGAAACCGCCTGGGAA 7372

Qy        301 ACCATCGAACACGCCGGCATCAACCCCCACACCCTCCACGGCACCCCCACCGGAGTCTTC 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7373 ACCATCGAACACGCCGGCATCAACCCCCACACCCTCCACGGCACCCCCACCGGAGTCTTC 7432

Qy        361 ACCGGCACCAACGGACAGGACTACGCACTTCGCGTGCACAACGCGGGCCAGTCAACCGAT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7433 ACCGGCACCAACGGACAGGACTACGCACTTCGCGTGCACAACGCGGGCCAGTCAACCGAT 7492

Qy        421 GGTTTCGCACTGACCGGAACCGCCGGCAGCGTCATCTCCGGTCGTATCTCGTACACGTTT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7493 GGTTTCGCACTGACCGGAACCGCCGGCAGCGTCATCTCCGGTCGTATCTCGTACACGTTT 7552

Qy        481 GGTTTTGAGGGTCCTGCGGTGTCGGTGGACACGGCTTGTTCCTCGTCGTTGGTGGCTTTG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7553 GGTTTTGAGGGTCCTGCGGTGTCGGTGGACACGGCTTGTTCCTCGTCGTTGGTGGCTTTG 7612

Qy        541 CATCTGGCCTGTCAGGCGTTGCGTGCGGGTGAGTGCTCGATGGCGCTTGCCGGGGGTGTG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7613 CATCTGGCCTGTCAGGCGTTGCGTGCGGGTGAGTGCTCGATGGCGCTTGCCGGGGGTGTG 7672

Qy        601 ACGGTGATGTCGTCTCCGGGTGCCTTCGTGGAGTTTTCGCGGCAGCGGGGTCTGGCCGCG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7673 ACGGTGATGTCGTCTCCGGGTGCCTTCGTGGAGTTTTCGCGGCAGCGGGGTCTGGCCGCG 7732

Qy        661 GACGGGCATTGCAAGGCGTTCTCGGCGGCGGCGGACGGGACCGGCTGGGGTGAGGGTGTG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db       7733 GACGGGCATTGCAAGGCGTTCTCGGCGGCGGCGGACGGGACCGGCTGGGGTGAGGGTGTG 7792

Qy        721 GGGATGCTGCTGGTGGAGCGGCTCTCCGACGCCCATCGCAACGGTCACCGTGTCCTGGCC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7793 GGGATGCTGCTGGTGGAGCGGCTCTCCGACGCCCATCGCAACGGTCACCGTGTCCTGGCC 7852

Qy        781 GTGGTGCGTGGCAGTGCGGTCAACCAGGACGGTGCGAGCAACGGTCTGACCGCGCCCAAC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7853 GTGGTGCGTGGCAGTGCGGTCAACCAGGACGGTGCGAGCAACGGTCTGACCGCGCCCAAC 7912

Qy        841 GGGCCGTCCCAGCAGCGTGTCATCCGCCAGGCCCTCGCCAACGCCGGCTTGTCGGCCGGT 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7913 GGGCCGTCCCAGCAGCGTGTCATCCGCCAGGCCCTCGCCAACGCCGGCTTGTCGGCCGGT 7972

Qy        901 GATGTCGACGCGGTGGAGGCCCACGGCACCGGCACCACTTTGGGCGACCCGATCGAGGCC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7973 GATGTCGACGCGGTGGAGGCCCACGGCACCGGCACCACTTTGGGCGACCCGATCGAGGCC 8032

Qy        961 CAGGCCCTCCTCGCGACCTACGGACAGGACCGTGCCGGCGAGGGGCCGCTGTGGCTGGGC 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8033 CAGGCCCTCCTCGCGACCTACGGACAGGACCGTGCCGGCGAGGGGCCGCTGTGGCTGGGC 8092

Qy       1021 TCGGTCAAGTCCAATGTCGGTCACACACAGGCTGCCGCGGGCGTCGCCGGGGTGATCAAG 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8093 TCGGTCAAGTCCAATGTCGGTCACACACAGGCTGCCGCGGGCGTCGCCGGGGTGATCAAG 8152

Qy       1081 ATGGTGATGGCGCTGCGGCATGGTCTGCTGCCGCGGACGTTGCATGTGGATGAGCCGTCG 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8153 ATGGTGATGGCGCTGCGGCATGGTCTGCTGCCGCGGACGTTGCATGTGGATGAGCCGTCG 8212

Qy       1141 CCGCATGTGGACTGGTCCGCGGGTGCGGTGCAGCTGCTGACGGAGACGGTGCCCTGGCCC 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8213 CCGCATGTGGACTGGTCCGCGGGTGCGGTGCAGCTGCTGACGGAGACGGTGCCCTGGCCC 8272

Qy       1201 GGCGGGGAGGGGCGGCTACGGCGGGCAGGAGTGTCATCATTCGGCGTCAGCGGCACCAAC 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8273 GGCGGGGAGGGGCGGCTACGGCGGGCAGGAGTGTCATCATTCGGCGTCAGCGGCACCAAC 8332

Qy       1261 GCCCACGTCATCCTCGAAGAAGCACCCGCC 1290
              ||||||||||||||||||||||||||||||
Db       8333 GCCCACGTCATCCTCGAAGAAGCACCCGCC 8362


RESULT 3
AAA92301
ID   AAA92301 standard; DNA; 30690 BP.
XX
AC   AAA92301;
XX
DT   10-JAN-2001  (first entry)
XX
DE   S. avermitilis avermectin aglycon synthase DNA aveAI SEQ ID NO:1.
XX
KW   Streptomyces avermitilis; avermectin aglycon synthase; biosynthesis;
KW   multifunctional enzyme; polyketide; avermectin; veterinary drug;
KW   agrochemical; ds.
XX
OS   Streptomyces avermitilis.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..11919
FT                   /*tag= a
FT                   /note= "avermectin aglycon synthase protein"
FT   CDS             11971..30690
FT                   /*tag= b
FT                   /note= "avermectin aglycon synthase protein"
XX
PN   WO200050605-A1.
XX
PD   31-AUG-2000.
XX
PF   23-FEB-2000; 2000WO-JP01041.
XX
```

```
PR   24-FEB-1999;   99JP-0046961.
XX
PA   (KITA ) KITASATO INST.
XX
PI   Omura S,  Ikeda H;
XX
DR   WPI; 2000-565458/52.
DR   P-PSDB; AAB23749, AAB23750.
XX
PT   Avermectin aglycone synthase DNA and proteins encoded by all or part of
PT   it for the production of avermectin and its derivatives for drug and
PT   agrochemical use  -
XX
PS   Claim 2; Page 66-134; 314pp; Japanese.
XX
CC   The present sequence represents DNA which encodes avermectin aglycon
CC   synthase proteins. Also described are: (1) polypeptides encoded by all
CC   or part of the DNA; (2) expression vectors containing the DNA; (3) host
CC   cells transformed by the vectors; (4) preparation of the polypeptides
CC   by culture of the transformants; (5) preparation of avermectin aglycon
CC   or its derivatives by culture of transformed avermectin-producing
CC   microorganisms; and (6) oligonucleotides of 5-60 bases in length
CC   containing sense or antisense sequences from the avermectin aglycon
CC   synthase DNA. The enzymes are useful for the production of modified
CC   forms of avermectin and of the intermediates in its biosynthesis, for
CC   use as drugs, veterinary drugs and agrochemicals.
XX
SQ   Sequence 30690 BP; 5356 A; 12454 C; 8617 G; 4263 T; 0 other;

  Query Match             100.0%;  Score 1290;  DB 21;  Length 30690;
  Best Local Similarity   100.0%;  Pred. No. 1.2e-206;
  Matches 1290;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GAGCCGATCGCCATCGTTGGCATGGCCTGTCGTTTCCCCGGCGGAGTGACCTCGGCGGAC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6256 GAGCCGATCGCCATCGTTGGCATGGCCTGTCGTTTCCCCGGCGGAGTGACCTCGGCGGAC 6315

Qy         61 GACTTCTGGGATCTGATCTCCTCCGAGCAGGACGCGATCGGCGGATTCCCCACCGACCGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6316 GACTTCTGGGATCTGATCTCCTCCGAGCAGGACGCGATCGGCGGATTCCCCACCGACCGC 6375

Qy        121 GGCTGGGACCTGGACACGCTCTACGACCCCGACCCCGACCACCCCGGCACCTGCTACACC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6376 GGCTGGGACCTGGACACGCTCTACGACCCCGACCCCGACCACCCCGGCACCTGCTACACC 6435

Qy        181 CGAAACGGCGGATTCCTCTACGACGCAGGCCACTTCGACGCCGAATTCTTCGGCATCAGC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6436 CGAAACGGCGGATTCCTCTACGACGCAGGCCACTTCGACGCCGAATTCTTCGGCATCAGC 6495

Qy        241 CCCCGCGAAGCCCTCGCCATGGACCCCCAGCAACGACTCCTCCTCGAAACCGCCTGGGAA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6496 CCCCGCGAAGCCCTCGCCATGGACCCCCAGCAACGACTCCTCCTCGAAACCGCCTGGGAA 6555

Qy        301 ACCATCGAACACGCCGGCATCAACCCCCACACCCTCCACGGCACCCCCACCGGAGTCTTC 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6556 ACCATCGAACACGCCGGCATCAACCCCCACACCCTCCACGGCACCCCCACCGGAGTCTTC 6615

Qy        361 ACCGGCACCAACGGACAGGACTACGCACTTCGCGTGCACAACGCGGGCCAGTCAACCGAT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6616 ACCGGCACCAACGGACAGGACTACGCACTTCGCGTGCACAACGCGGGCCAGTCAACCGAT 6675

Qy        421 GGTTTCGCACTGACCGGAACCGCCGGCAGCGTCATCTCCGGTCGTATCTCGTACACGTTT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6676 GGTTTCGCACTGACCGGAACCGCCGGCAGCGTCATCTCCGGTCGTATCTCGTACACGTTT 6735

Qy        481 GGTTTTGAGGGTCCTGCGGTGTCGGTGGACACGGCTTGTTCCTCGTCGTTGGTGGCTTTG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6736 GGTTTTGAGGGTCCTGCGGTGTCGGTGGACACGGCTTGTTCCTCGTCGTTGGTGGCTTTG 6795

Qy        541 CATCTGGCCTGTCAGGCGTTGCGTGCGGGTGAGTGCTCGATGGCGCTTGCCGGGGGTGTG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6796 CATCTGGCCTGTCAGGCGTTGCGTGCGGGTGAGTGCTCGATGGCGCTTGCCGGGGGTGTG 6855

Qy        601 ACGGTGATGTCGTCTCCGGGTGCCTTCGTGGAGTTTTCGCGGCAGCGGGGTCTGGCCGCG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6856 ACGGTGATGTCGTCTCCGGGTGCCTTCGTGGAGTTTTCGCGGCAGCGGGGTCTGGCCGCG 6915

Qy        661 GACGGGCATTGCAAGGCGTTCTCGGCGGCGGCGGACGGGACCGGCTGGGGTGAGGGTGTG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6916 GACGGGCATTGCAAGGCGTTCTCGGCGGCGGCGGACGGGACCGGCTGGGGTGAGGGTGTG 6975

Qy        721 GGGATGCTGCTGGTGGAGCGGCTCTCCGACGCCCATCGCAACGGTCACCGTGTCCTGGCC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6976 GGGATGCTGCTGGTGGAGCGGCTCTCCGACGCCCATCGCAACGGTCACCGTGTCCTGGCC 7035

Qy        781 GTGGTGCGTGGCAGTGCGGTCAACCAGGACGGTGCGAGCAACGGTCTGACCGCGCCCAAC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7036 GTGGTGCGTGGCAGTGCGGTCAACCAGGACGGTGCGAGCAACGGTCTGACCGCGCCCAAC 7095

Qy        841 GGGCCGTCCCAGCAGCGTGTCATCCGCCAGGCCCTCGCCAACGCCGGCTTGTCGGCCGGT 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7096 GGGCCGTCCCAGCAGCGTGTCATCCGCCAGGCCCTCGCCAACGCCGGCTTGTCGGCCGGT 7155

Qy        901 GATGTCGACGCGGTGGAGGCCCACGGCACCGGCACCACTTTGGGCGACCCGATCGAGGCC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7156 GATGTCGACGCGGTGGAGGCCCACGGCACCGGCACCACTTTGGGCGACCCGATCGAGGCC 7215

Qy        961 CAGGCCCTCCTCGCGACCTACGGACAGGACCGTGCCGGCGAGGGGCCGCTGTGGCTGGGC 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7216 CAGGCCCTCCTCGCGACCTACGGACAGGACCGTGCCGGCGAGGGGCCGCTGTGGCTGGGC 7275

Qy       1021 TCGGTCAAGTCCAATGTCGGTCACACACAGGCTGCCGCGGGCGTCGCCGGGGTGATCAAG 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7276 TCGGTCAAGTCCAATGTCGGTCACACACAGGCTGCCGCGGGCGTCGCCGGGGTGATCAAG 7335

Qy       1081 ATGGTGATGGCGCTGCGGCATGGTCTGCTGCCGCGGACGTTGCATGTGGATGAGCCGTCG 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7336 ATGGTGATGGCGCTGCGGCATGGTCTGCTGCCGCGGACGTTGCATGTGGATGAGCCGTCG 7395

Qy       1141 CCGCATGTGGACTGGTCCGCGGGTGCGGTGCAGCTGCTGACGGAGACGGTGCCCTGGCCC 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7396 CCGCATGTGGACTGGTCCGCGGGTGCGGTGCAGCTGCTGACGGAGACGGTGCCCTGGCCC 7455

Qy       1201 GGCGGGGAGGGGCGGCTACGGCGGGCAGGAGTGTCATCATTCGGCGTCAGCGGCACCAAC 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7456 GGCGGGGAGGGGCGGCTACGGCGGGCAGGAGTGTCATCATTCGGCGTCAGCGGCACCAAC 7515

Qy       1261 GCCCACGTCATCCTCGAAGAAGCACCCGCC 1290
              |||||||||||||||||||||||||||||
Db       7516 GCCCACGTCATCCTCGAAGAAGCACCCGCC 7545


RESULT 4
AAH79277
ID   AAH79277 standard; DNA; 30690 BP.
XX
AC   AAH79277;
XX
DT   04-DEC-2001  (first entry)
XX
DE   Streptomyces avermitilis coding sequences SEQ ID NO: 1.
XX
KW   Avermectin aglycone synthase; AAS; avermectin derivative;
KW   drug production; vetinary drug; pesticide; ds.
XX
OS   Streptomyces avermitilis.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..11919
FT                   /*tag=  a
FT                   /product= "AAG65264"
FT                   /partial
FT   CDS             11971..30690
FT                   /*tag=  b
FT                   /product= "AAG65265"
XX
```

```
PN   WO200162939-A1.
XX
PD   30-AUG-2001.
XX
PF   23-FEB-2001; 2001WO-JP01381.
XX
PR   24-FEB-2000; 2000JP-0047405.
XX
PA   (KYOW ) KYOWA HAKKO KOGYO KK.
PA   (KITA ) KITASATO INST.
XX
PI   Endo H,  Yamaguchi H,  Kanda Y,  Hashimoto S,  Omura S,  Ikeda H;
XX
DR   WPI; 2001-582053/65.
DR   P-PSDB; AAG65264, AAG65265.
XX
PT   New modified avermectin aglycone synthase derived from Streptomyces
PT   avermectilis used in production of 22,23-dihydroavermectin B1a used in
PT   drugs and pesticides -
XX
PS   Example 2; Page 58-123; 257pp; Japanese.
XX
CC   The present invention relates to the production of modified derivatives
CC   of avermectin aglycone synthase (AAS) derived from Streptomyces
CC   avermitilis. The activity of an acyl carrier protein (ACP),
CC   beta-ketoacyl carrier protein synthase (KS), acyltransferase (AT),
CC   beta-ketoacyl carrier protein reductase (KR), dehydratase (DH), enoyl
CC   reductase (ER) and/or thioesterase (TE) domain may be reduced or
CC   suppressed. The process can be used in the production of drugs, vetinary
CC   drugs and pesticides. The present sequence is a fragment of the S.
CC   avermitilis genome.
XX
SQ   Sequence 30690 BP; 5355 A; 12455 C; 8617 G; 4263 T; 0 other;

  Query Match             100.0%;  Score 1290;  DB 22;  Length 30690;
  Best Local Similarity   100.0%;  Pred. No. 1.2e-206;
  Matches 1290;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GAGCCGATCGCCATCGTTGGCATGGCCTGTCGTTTCCCCGGCGGAGTGACCTCGGCGGAC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6256 GAGCCGATCGCCATCGTTGGCATGGCCTGTCGTTTCCCCGGCGGAGTGACCTCGGCGGAC 6315

Qy         61 GACTTCTGGGATCTGATCTCCTCCGAGCAGGACGCGATCGGCGGATTCCCCACCGACCGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6316 GACTTCTGGGATCTGATCTCCTCCGAGCAGGACGCGATCGGCGGATTCCCCACCGACCGC 6375

Qy        121 GGCTGGGACCTGGACACGCTCTACGACCCCGACCCCGACCACCCCGGCACCTGCTACACC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6376 GGCTGGGACCTGGACACGCTCTACGACCCCGACCCCGACCACCCCGGCACCTGCTACACC 6435

Qy        181 CGAAACGGCGGATTCCTCTACGACGCAGGCCACTTCGACGCCGAATTCTTCGGCATCAGC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6436 CGAAACGGCGGATTCCTCTACGACGCAGGCCACTTCGACGCCGAATTCTTCGGCATCAGC 6495

Qy        241 CCCCGCGAAGCCCTCGCCATGGACCCCCAGCAACGACTCCTCCTCGAAACCGCCTGGGAA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6496 CCCCGCGAAGCCCTCGCCATGGACCCCCAGCAACGACTCCTCCTCGAAACCGCCTGGGAA 6555

Qy        301 ACCATCGAACACGCCGGCATCAACCCCCACACCCTCCACGGCACCCCCACCGGAGTCTTC 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6556 ACCATCGAACACGCCGGCATCAACCCCCACACCCTCCACGGCACCCCCACCGGAGTCTTC 6615

Qy        361 ACCGGCACCAACGGACAGGACTACGCACTTCGCGTGCACAACGCGGGCCAGTCAACCGAT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6616 ACCGGCACCAACGGACAGGACTACGCACTTCGCGTGCACAACGCGGGCCAGTCAACCGAT 6675

Qy        421 GGTTTCGCACTGACCGGAACCGCCGGCAGCGTCATCTCCGGTCGTATCTCGTACACGTTT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6676 GGTTTCGCACTGACCGGAACCGCCGGCAGCGTCATCTCCGGTCGTATCTCGTACACGTTT 6735

Qy        481 GGTTTTGAGGGTCCTGCGGTGTCGGTGGACACGGCTTGTTCCTCGTCGTTGGTGGCTTTG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6736 GGTTTTGAGGGTCCTGCGGTGTCGGTGGACACGGCTTGTTCCTCGTCGTTGGTGGCTTTG 6795

Qy        541 CATCTGGCCTGTCAGGCGTTGCGTGCGGGTGAGTGCTCGATGGCGCTTGCCGGGGGTGTG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6796 CATCTGGCCTGTCAGGCGTTGCGTGCGGGTGAGTGCTCGATGGCGCTTGCCGGGGGTGTG 6855

Qy        601 ACGGTGATGTCGTCTCCGGGTGCCTTCGTGGAGTTTTCGCGGCAGCGGGGTCTGGCCGCG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6856 ACGGTGATGTCGTCTCCGGGTGCCTTCGTGGAGTTTTCGCGGCAGCGGGGTCTGGCCGCG 6915

Qy        661 GACGGGCATTGCAAGGCGTTCTCGGCGGCGGCGGACGGGACCGGCTGGGGTGAGGGTGTG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6916 GACGGGCATTGCAAGGCGTTCTCGGCGGCGGCGGACGGGACCGGCTGGGGTGAGGGTGTG 6975

Qy        721 GGGATGCTGCTGGTGGAGCGGCTCTCCGACGCCCATCGCAACGGTCACCGTGTCCTGGCC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6976 GGGATGCTGCTGGTGGAGCGGCTCTCCGACGCCCATCGCAACGGTCACCGTGTCCTGGCC 7035

Qy        781 GTGGTGCGTGGCAGTGCGGTCAACCAGGACGGTGCGAGCAACGGTCTGACCGCGCCCAAC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7036 GTGGTGCGTGGCAGTGCGGTCAACCAGGACGGTGCGAGCAACGGTCTGACCGCGCCCAAC 7095

Qy        841 GGGCCGTCCCAGCAGCGTGTCATCCGCCAGGCCCTCGCCAACGCCGGCTTGTCGGCCGGT 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7096 GGGCCGTCCCAGCAGCGTGTCATCCGCCAGGCCCTCGCCAACGCCGGCTTGTCGGCCGGT 7155

Qy        901 GATGTCGACGCGGTGGAGGCCCACGGCACCGGCACCACTTTGGGCGACCCGATCGAGGCC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7156 GATGTCGACGCGGTGGAGGCCCACGGCACCGGCACCACTTTGGGCGACCCGATCGAGGCC 7215

Qy        961 CAGGCCCTCCTCGCGACCTACGGACAGGACCGTGCCGGCGAGGGGCCGCTGTGGCTGGGC 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7216 CAGGCCCTCCTCGCGACCTACGGACAGGACCGTGCCGGCGAGGGGCCGCTGTGGCTGGGC 7275

Qy       1021 TCGGTCAAGTCCAATGTCGGTCACACACAGGCTGCCGCGGGCGTCGCCGGGGTGATCAAG 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7276 TCGGTCAAGTCCAATGTCGGTCACACACAGGCTGCCGCGGGCGTCGCCGGGGTGATCAAG 7335

Qy       1081 ATGGTGATGGCGCTGCGGCATGGTCTGCTGCCGCGGACGTTGCATGTGGATGAGCCGTCG 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7336 ATGGTGATGGCGCTGCGGCATGGTCTGCTGCCGCGGACGTTGCATGTGGATGAGCCGTCG 7395

Qy       1141 CCGCATGTGGACTGGTCCGCGGGTGCGGTGCAGCTGCTGACGGAGACGGTGCCCTGGCCC 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7396 CCGCATGTGGACTGGTCCGCGGGTGCGGTGCAGCTGCTGACGGAGACGGTGCCCTGGCCC 7455

Qy       1201 GGCGGGGAGGGGCGGCTACGGCGGGCAGGAGTGTCATCATTCGGCGTCAGCGGCACCAAC 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7456 GGCGGGGAGGGGCGGCTACGGCGGGCAGGAGTGTCATCATTCGGCGTCAGCGGCACCAAC 7515

Qy       1261 GCCCACGTCATCCTCGAAGAAGCACCCGCC 1290
              |||||||||||||||||||||||||||||
Db       7516 GCCCACGTCATCCTCGAAGAAGCACCCGCC 7545


RESULT 5
AAA92302
ID   AAA92302 standard; DNA; 31422 BP.
XX
AC   AAA92302;
XX
DT   10-JAN-2001  (first entry)
XX
DE   S. avermitilis avermectin aglycon synthase DNA aveAII SEQ ID NO:2.
XX
KW   Streptomyces avermitilis; avermectin aglycon synthase; biosynthesis;
KW   multifunctional enzyme; polyketide; avermectin; veterinary drug;
KW   agrochemical; ds.
XX
OS   Streptomyces avermitilis.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..14646
FT                   /*tag=  a
```

```
FT                   /note= "avermectin aglycon synthase protein"
FT   CDS             14824..31422
FT                   /*tag=  b
FT                   /note= "avermectin aglycon synthase protein"
XX
PN   WO200050605-A1.
XX
PD   31-AUG-2000.
XX
PF   23-FEB-2000; 2000WO-JP01041.
XX
PR   24-FEB-1999;   99JP-0046961.
XX
PA   (KITA ) KITASATO INST.
XX
PI   Omura S,  Ikeda H;
XX
DR   WPI; 2000-565458/52.
DR   P-PSDB; AAB23751, AAB23752.
XX
PT   Avermectin aglycone synthase DNA and proteins encoded by all or part of
PT   it for the production of avermectin and its derivatives for drug and
PT   agrochemical use  -
XX
PS   Claim 2; Page 134-203; 314pp; Japanese.
XX
CC   The present sequence represents DNA which encodes avermectin aglycon
CC   synthase proteins. Also described are: (1) polypeptides encoded by all
CC   or part of the DNA; (2) expression vectors containing the DNA; (3) host
CC   cells transformed by the vectors; (4) preparation of the polypeptides
CC   by culture of the transformants; (5) preparation of avermectin aglycon
CC   or its derivatives by culture of transformed avermectin-producing
CC   microorganisms; and (6) oligonucleotides of 5-60 bases in length
CC   containing sense or antisense sequences from the avermectin aglycon
CC   synthase DNA. The enzymes are useful for the production of modified
CC   forms of avermectin and of the intermediates in its biosynthesis, for
CC   use as drugs, veterinary drugs and agrochemicals.
XX
SQ   Sequence 31422 BP; 4136 A; 10238 C; 11677 G; 5371 T; 0 other;

  Query Match             80.0%;  Score 1032.2;  DB 21;  Length 31422;
  Best Local Similarity   87.9%;  Pred. No. 1.1e-163;
  Matches 1137;  Conservative    0;  Mismatches  153;  Indels    3;  Gaps    1;

Qy          1 GAGCCGATCGCCATCGTTGGCATGGCCTGTCGTTTCCCCGGCGGAGTGACCTCGGCGGAC 60
              ||||| |||| ||| |||||||||| || || | ||||| || || ||||| || ||
Db      20413 GAGCCCGTCGCGATCATTGGCATGGCGTGCCGCTATCCCGGAGGCGTCACCTCAGCCGAG 20472

Qy         61 GACTTCTGGGATCTGATCTCCTCCGAGCAGGACGCGATCGGCGGATTCCCCACCGACCGC 120
              ||  | ||||| ||| || | || | |  |||| || ||||||  || || || |||||
Db      20473 GAGCTGTGGGAACTGCTCGCATCGGGGAGGGACACGGTCGGCGAGTTTCCGACGGACCGT 20532

Qy        121 GGCTGGGACCTGGACACGCTCTACGACCCCGACCCCGACCACCCCGGCACCTGCTACACC 180
              || ||||||||||  | || | ||| || || || |  |  || |||||||  ||||||
Db      20533 GGGTGGGACCTGGAAGCACTGTTCGATCCGGAACCGGGTCGGCCGGGCACCTCGTACACC 20592

Qy        181 CGAAACGGCGGATTCCTCTACGACGCAGGCCACTTCGACGCCGAATTCTTCGGCATCAGC 240
              ||    ||  | ||||||||||||||| ||  | ||||||||||  |||||||| |||||
Db      20593 CGCTGTGGGAGTTTCCTCTACGACGCGGGGGAGTTCGACGCCGGCTTCTTCGGGATCAGT 20652

Qy        241 CCCCGCGAAGCCCTCGCCATGGACCCCCAGCAACGACTCCTCCTCGAAACCGCCTGGGAA 300
              || || || || || || |||||||| ||||| ||| | || || ||  || | |||||
Db      20653 CCGCGTGAGGCACTGGCGATGGACCCGCAGCAGCGATTGCTGCTGGAGGCCTCATGGGAG 20712

Qy        301 ACCATCGAACACGCCGGCATCAACCCCCACACCCTCCACGGCACCCCCACCGGAGTCTTC 360
              ||||  || || || || ||  ||||    ||| | | ||| | ||  || || || |||
Db      20713 GCCATGGAGCAGGCAGGTATTGACCCTACGACCGTACGCGGGAGCCAGACAGGCGTGTTC 20772

Qy        361 ACCGGCACCAACGGACAGGACTACGCACTTCGCGTGCAC---AACGCGGGCCAGTCAACC 417
              | |||  ||    |||| ||| | ||   |  |||||  ||||| |        |||
Db      20773 GCGGGCCTCATTCCGCAGGCCTATGGACCCAGGCTGCACGAAAACGCCGCAGCCGACACC 20832

Qy        418 GATGGTTTCGCACTGACCGGAACCGCCGGCAGCGTCATCTCCGGTCGTATCTCGTACACG 477
              || || |  | |||||||| ||  |||| || ||   |||||||||||||||||||||||
Db      20833 GAGGGCTATGTCCTGACCGGCACATCCGGGAGTGTGGCCTCCGGTCGTATCTCGTACACG 20892

Qy        478 TTTGGTTTTGAGGGTCCTGCGGTGTCGGTGGACACGGCTTGTTCCTCGTCGTTGGTGGCT 537
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      20893 TTTGGTTTTGAGGGTCCTGCGGTGTCGGTGGACACGGCTTGTTCCTCGTCGTTGGTGGCT 20952

Qy        538 TTGCATCTGGCCTGTCAGGCGTTGCGTGCGGGTGAGTGCTCGATGGCGCTTGCCGGGGGT 597
              || ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      20953 TTACATCTGGCCTGTCAGGCGTTGCGTGCGGGTGAGTGCTCGATGGCGCTTGCCGGGGGT 21012

Qy        598 GTGACGGTGATGTCGTCTCCGGGTGCCTTCGTGGAGTTTTCGCGGCAGCGGGGTCTGGCC 657
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      21013 GTGACGGTGATGTCGTCTCCGGGTGCCTTCGTGGAGTTTTCGCGGCAGCGGGGTCTGGCC 21072

Qy        658 GCGGACGGGCATTGCAAGGCGTTCTCGGCGGCGGCGGACGGGACCGGCTGGGGTGAGGGT 717
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      21073 GCGGACGGGCATTGCAAGGCGTTCTCGGCGGCGGCGGACGGGACCGGCTGGGGTGAGGGT 21132

Qy        718 GTGGGGATGCTGCTGGTGGAGCGGCTCTCCGACGCCCATCGCAACGGTCACCGTGTCCTG 777
              |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db      21133 GTGGGGATGCTGCTGGTGGAGCGGCTCTCCGACGCCCGTCGCAACGGTCACCGTGTCCTG 21192

Qy        778 GCCGTGGTGCGTGGCAGTGCGGTCAACCAGGACGGTGCGAGCAACGGTCTGACCGCGCCC 837
              |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db      21193 GCCGTGGTGCGTGGCAGTGCGGTCAACCAGGACGGTGCGAGCAACGGGCTGACCGCGCCC 21252

Qy        838 AACGGGCCGTCCCAGCAGCGTGTCATCCGCCAGGCCCTCGCCAACGCCGGCTTGTCGGCC 897
              ||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db      21253 AACGGGCCCTCCCAGCAGCGTGTCATCCGCCAGGCCCTCGCCAACGCCGGCTTGTCGGCC 21312

Qy        898 GGTGATGTCGACGCGGTGGAGGCCCACGGCACCGGCACCACTTTGGGCGACCCGATCGAG 957
              |||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db      21313 GGTGATGTCGATGCGGTGGAGGCCCACGGCACCGGCACCACTTTGGGCGACCCGATCGAG 21372

Qy        958 GCCCAGGCCCTCCTCGCGACCTACGGACAGGACCGTGCCGGCGAGGGGCCGCTGTGGCTG 1017
              |||||||||||||| |||||||||| ||||||||||||||||||||||||||||||||||
Db      21373 GCCCAGGCCCTCCTTGCGACCTACGGGCAGGACCGTGCCGGCGAGGGGCCGCTGTGGCTG 21432

Qy       1018 GGCTCGGTCAAGTCCAATGTCGGTCACACACAGGCTGCCGCGGGCGTCGCCGGGGTGATC 1077
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      21433 GGCTCGGTCAAGTCCAATGTCGGTCACACACAGGCTGCCGCGGGCGTCGCCGGGGTGATC 21492

Qy       1078 AAGATGGTGATGGCGCTGCGGCATGGTCTGCTGCCGCGGACGTTGCATGTGGATGAGCCG 1137
              |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db      21493 AAGATGGTGATGGCGCTGCGGAATGGTCTGCTGCCGCGGACGTTGCATGTGGATGAGCCG 21552

Qy       1138 TCGCCGCATGTGGACTGGTCCGCGGGTGCGGTGCAGCTGCTGACGGAGACGGTGCCCTGG 1197
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      21553 TCGCCGCATGTGGACTGGTCCGCGGGTGCGGTGCAGCTGCTGACGGAGACGGTGCCCTGG 21612

Qy       1198 CCCGGCGGGGAGGGGCGGCTACGGCGGGCAGGAGTGTCATCATTCGGCGTCAGCGGCACC 1257
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      21613 CCCGGCGGGGAGGGGCGGCTACGGCGGGCAGGAGTGTCATCATTCGGCGTCAGCGGCACC 21672

Qy       1258 AACGCCCACGTCATCCTCGAAGAAGCACCCGCC 1290
              |||||||||||||||||||||||||||||||||
Db      21673 AACGCCCACGTCATCCTCGAAGAAGCACCCGCC 21705


RESULT 6
AAH79278
ID   AAH79278 standard; DNA; 31422 BP.
XX
AC   AAH79278;
XX
DT   04-DEC-2001  (first entry)
XX
DE   Streptomyces avermitilis coding sequences SEQ ID NO: 2.
XX
KW   Avermectin aglycone synthase; AAS; avermectin derivative;
KW   drug production; vetinary drug; pesticide; ds.
XX
```

```
OS   Streptomyces avermitilis.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..14646
FT                   /*tag= a
FT                   /product= "AAG65266"
FT   CDS             14824..31422
FT                   /*tag= b
FT                   /product= "AAG65267"
XX
PN   WO200162939-A1.
XX
PD   30-AUG-2001.
XX
PF   23-FEB-2001; 2001WO-JP01381.
XX
PR   24-FEB-2000; 2000JP-0047405.
XX
PA   (KYOW ) KYOWA HAKKO KOGYO KK.
PA   (KITA ) KITASATO INST.
XX
PI   Endo H,  Yamaguchi H,  Kanda Y,  Hashimoto S,  Omura S,  Ikeda H;
XX
DR   WPI; 2001-582053/65.
DR   P-PSDB; AAG65266, AAG65267.
XX
PT   New modified avermectin aglycone synthase derived from Streptomyces
PT   avermectilis used in production of 22,23-dihydroavermectin B1a used in
PT   drugs and pesticides -
XX
PS   Disclosure; Page 103-149; 257pp; Japanese.
XX
CC   The present invention relates to the production of modified derivatives
CC   of avermectin aglycone synthase (AAS) derived from Streptomyces
CC   avermitilis. The activity of an acyl carrier protein (ACP),
CC   beta-ketoacyl carrier protein synthase (KS), acyltransferase (AT),
CC   beta-ketoacyl carrier protein reductase (KR), dehydratase (DH), enoyl
CC   reductase (ER) and/or thioesterase (TE) domain may be reduced or
CC   suppressed. The process can be used in the production of drugs, vetinary
CC   drugs and pesticides. The present sequence is a fragment of the S.
CC   avermitilis genome.
XX
SQ   Sequence 31422 BP; 4136 A; 10237 C; 11677 G; 5372 T; 0 other;

  Query Match             80.0%;  Score 1032.2;  DB 22;  Length 31422;
  Best Local Similarity   87.9%;  Pred. No. 1.1e-163;
  Matches 1137;  Conservative    0;  Mismatches  153;  Indels    3;  Gaps    1;

Qy          1 GAGCCGATCGCCATCGTTGGCATGGCCTGTCGTTTCCCCGGCGGAGTGACCTCGGCGGAC 60
              |||||  ||||  ||| |||||||||| || || |  |||||  || ||  ||||| || ||
Db      20413 GAGCCCGTCGCGATCATTGGCATGGCGTGCCGCTATCCCGGAGGCGTCACCTCAGCCGAG 20472

Qy         61 GACTTCTGGGATCTGATCTCCTCCGAGCAGGACGCGATCGGCGGATTCCCCACCGACCGC 120
              ||  | |||||  ||| || | || | |  |||| || ||||||  || || || |||||
Db      20473 GAGCTGTGGGAACTGCTCGCATCGGGGAGGGACACGGTCGGCGAGTTTCCGACGGACCGT 20532

Qy        121 GGCTGGGACCTGGACACGCTCTACGACCCCGACCCCGACCACCCCGGCACCTGCTACACC 180
              || ||||||||||  | || | ||| || || || |  |  || |||||||  ||||||
Db      20533 GGGTGGGACCTGGAAGCACTGTTCGATCCGGAACCGGGTCGGCCGGGCACCTCGTACACC 20592

Qy        181 CGAAACGGCGGATTCCTCTACGACGCAGGCCACTTCGACGCCGAATTCTTCGGCATCAGC 240
              ||    || .| |||||||||||||| ||  | ||||||||||  ||||||||  |||||
Db      20593 CGCTGTGGGAGTTTCCTCTACGACGCGGGGGAGTTCGACGCCGGCTTCTTCGGGATCAGT 20652

Qy        241 CCCCGCGAAGCCCTCGCCATGGACCCCCAGCAACGACTCCTCCTCGAAACCGCCTGGGAA 300
              || || || || || || |||||||| |||| ||| | || || ||  || | |||||
Db      20653 CCGCGTGAGGCACTGGCGATGGACCCGCAGCAGCGATTGCTGCTGGAGGCCTCATGGGAG 20712

Qy        301 ACCATCGAACACGCCGGCATCAACCCCCACACCCTCCACGGCACCCCCACCGGAGTCTTC 360
               |||| || || || || ||  ||||    ||| | | ||| | ||  || || || |||
Db      20713 GCCATGGAGCAGGCAGGTATTGACCCTACGACCGTACGCGGGAGCCAGACAGGCGTGTTC 20772

Qy        361 ACCGGCACCAACGGACAGGACTACGCACTTCGCGTGCAC---AACGCGGGCCAGTCAACC 417
                | |||  ||     |||| ||| | ||  |  |||||  |||||  |        |||
Db      20773 GCGGGCCTCATTCCGCAGGCCTATGGACCCAGGCTGCACGAAAACGCCGCAGCCGACACC 20832

Qy        418 GATGGTTTCGCACTGACCGGAACCGCCGGCAGCGTCATCTCCGGTCGTATCTCGTACACG 477
              || || |  | |||||||| ||  |||| || ||   ||||||||||||||||||||
Db      20833 GAGGGCTATGTCCTGACCGGCACATCCGGGAGTGTGGCCTCCGGTCGTATCTCGTACACG 20892

Qy        478 TTTGGTTTTGAGGGTCCTGCGGTGTCGGTGGACACGGCTTGTTCCTCGTCGTTGGTGGCT 537
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      20893 TTTGGTTTTGAGGGTCCTGCGGTGTCGGTGGACACGGCTTGTTCCTCGTCGTTGGTGGCT 20952

Qy        538 TTGCATCTGGCCTGTCAGGCGTTGCGTGCGGGTGAGTGCTCGATGGCGCTTGCCGGGGGT 597
              || ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      20953 TTACATCTGGCCTGTCAGGCGTTGCGTGCGGGTGAGTGCTCGATGGCGCTTGCCGGGGGT 21012

Qy        598 GTGACGGTGATGTCGTCTCCGGGTGCCTTCGTGGAGTTTTCGCGGCAGCGGGGTCTGGCC 657
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      21013 GTGACGGTGATGTCGTCTCCGGGTGCCTTCGTGGAGTTTTCGCGGCAGCGGGGTCTGGCC 21072

Qy        658 GCGGACGGGCATTGCAAGGCGTTCTCGGCGGCGGCGGACGGGACCGGCTGGGGTGAGGGT 717
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      21073 GCGGACGGGCATTGCAAGGCGTTCTCGGCGGCGGCGGACGGGACCGGCTGGGGTGAGGGT 21132

Qy        718 GTGGGGATGCTGCTGGTGGAGCGGCTCTCCGACGCCCATCGCAACGGTCACCGTGTCCTG 777
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      21133 GTGGGGATGCTGCTGGTGGAGCGGCTCTCCGACGCCCGTCGCAACGGTCACCGTGTCCTG 21192

Qy        778 GCCGTGGTGCGTGGCAGTGCGGTCAACCAGGACGGTGCGAGCAACGGTCTGACCGCGCCC 837
              |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db      21193 GCCGTGGTGCGTGGCAGTGCGGTCAACCAGGACGGTGCGAGCAACGGGCTGACCGCGCCC 21252

Qy        838 AACGGGCCGTCCCAGCAGCGTGTCATCCGCCAGGCCCTCGCCAACGCCGGCTTGTCGGCC 897
              |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      21253 AACGGGCCCTCCCAGCAGCGTGTCATCCGCCAGGCCCTCGCCAACGCCGGCTTGTCGGCC 21312

Qy        898 GGTGATGTCGACGCGGTGGAGGCCCACGGCACCGGCACCACTTTGGGCGACCCGATCGAG 957
              |||||||||| |||||||||||||||||||||||||||||||||||||||||||||||||
Db      21313 GGTGATGTCGATGCGGTGGAGGCCCACGGCACCGGCACCACTTTGGGCGACCCGATCGAG 21372

Qy        958 GCCCAGGCCCTCCTCGCGACCTACGGACAGGACCGTGCCGGCGAGGGGCCGCTGTGGCTG 1017
              |||||||||||||| |||||||||| ||||||||||||||||||||||||||||||||||
Db      21373 GCCCAGGCCCTCCTTGCGACCTACGGGCAGGACCGTGCCGGCGAGGGGCCGCTGTGGCTG 21432

Qy       1018 GGCTCGGTCAAGTCCAATGTCGGTCACACACAGGCTGCCGCGGGCGTCGCCGGGGTGATC 1077
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      21433 GGCTCGGTCAAGTCCAATGTCGGTCACACACAGGCTGCCGCGGGCGTCGCCGGGGTGATC 21492

Qy       1078 AAGATGGTGATGGCGCTGCGGCATGGTCTGCTGCCGCGGACGTTGCATGTGGATGAGCCG 1137
              |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db      21493 AAGATGGTGATGGCGCTGCGGAATGGTCTGCTGCCGCGGACGTTGCATGTGGATGAGCCG 21552

Qy       1138 TCGCCGCATGTGGACTGGTCCGCGGGTGCGGTGCAGCTGCTGACGGAGACGGTGCCCTGG 1197
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      21553 TCGCCGCATGTGGACTGGTCCGCGGGTGCGGTGCAGCTGCTGACGGAGACGGTGCCCTGG 21612

Qy       1198 CCCGGCGGGGAGGGGCGGCTACGGCGGGCAGGAGTGTCATCATTCGGCGTCAGCGGCACC 1257
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      21613 CCCGGCGGGGAGGGGCGGCTACGGCGGGCAGGAGTGTCATCATTCGGCGTCAGCGGCACC 21672

Qy       1258 AACGCCCACGTCATCCTCGAAGAAGCACCCGCC 1290
              |||||||||||||||||||||||||||||||||
Db      21673 AACGCCCACGTCATCCTCGAAGAAGCACCCGCC 21705


RESULT 7
AAT78508
ID   AAT78508 standard; DNA; 44377 BP.
XX
AC   AAT78508;
XX
DT   26-FEB-1998  (first entry)
XX
DE   Platenolide synthase gene cluster.
```

```
XX
KW   Platenolide synthase gene cluster; platenolide production; srmG gene;
KW   multi-functional protein; macrolide antibiotic; spiramycin; ss.
XX
OS   Streptomyces ambofaciens.
XX
FH   Key             Location/Qualifiers
FT   CDS             350..14002
FT                   /*tag=  a
FT                   /transl_except= (pos:350..352, aa:Met)
FT                   /note= "ORF1 encodes protein shown in AAW23716"
FT   CDS             14046..20036
FT                   /*tag=  b
FT                   /note= "ORF2 encodes protein shown in AAW23717"
FT   CDS             20110..31284
FT                   /*tag=  c
FT                   /transl_except= (pos:20111..20113, aa:Met)
FT                   /note= "ORF3 encodes protein shown in AAW23718"
FT   CDS             31329..36071
FT                   /*tag=  d
FT                   /note= "ORF4 encodes protein shown in AAW23719"
FT   CDS             36155..41830
FT                   /*tag=  e
FT                   /note= "ORF5 encodes protein shown in AAW23720"
XX
PN   EP791656-A2.
XX
PD   27-AUG-1997.
XX
PF   19-FEB-1997;   97EP-0301066.
XX
PR   22-FEB-1996;   96US-0012050.
XX
PA   (ELIL ) LILLY & CO ELI.
XX
PI   Burgett SG,  Kuhstoss SA,  Rao RN,  Richardson MA;
PI   Rosteck PR;
XX
DR   WPI; 1997-418047/39.
DR   P-PSDB; AAW23716-W23720.
XX
PT   DNA encoding Streptomyces ambofaciens platenolide synthase domain -
PT   for production of spiramycin-related polyketide antibiotics
XX
PS   Claim 9; Pages 8-33; 81pp; English.
XX
CC   This sequence represents the platenolide synthase gene cluster of the
CC   invention. This sequence is referred to as the srmG gene, and was
CC   isolated from Streptomyces ambofaciens. This sequence encodes the
CC   multi-functional proteins which direct the synthesis of the polyketide
CC   platenolide. Platenolide is the basic building block of the macrolide
CC   antibiotic spiramycin. The DNA can be used to produce compounds
CC   exhibiting antibiotic activity based on the platenolide structure,
CC   including specifically the macrolide antibiotic spiramycin and spiramycin
CC   analogues and derivatives. Modifications of the platenolide synthase DNA
CC   sequence can be made so as to change the number and type of carboxylic
CC   acids incorporated into the growing polyketide chain and to change the
CC   kind of post-condensation processing that is conducted.
XX
SQ   Sequence 44377 BP; 4965 A; 15552 C; 17381 G; 6479 T; 0 other;

  Query Match             56.8%;  Score 732.2;  DB 18;  Length 44377;
  Best Local Similarity   73.0%;  Pred. No. 1e-113;
  Matches  941;  Conservative    0;  Mismatches  348;  Indels    0;  Gaps    0;

Qy          1 GAGCCGATCGCCATCGTTGGCATGGCCTGTCGTTTCCCCGGCGGAGTGACCTCGGCGGAC 60
              |||||||||||| ||||| |||||||| || |||| ||| ||||| ||| | ||| |||||
Db      14148 GAGCCGATCGCGATCGTGGGCATGGCGTGCCGTTACCCGGGCGGTGTGGCGTCGCCGGAC 14207

Qy         61 GACTTCTGGGATCTGATCTCCTCCGAGCAGGACGCGATCGGCGGATTCCCCACCGACCGC 120
              ||| | ||||| ||| |   |  |||    ||||||| ||  ||  ||||||  ||||||
Db      14208 GACCTGTGGGACCTGGTGGCAGCCGGTACGGACGCGGTCTCCGCGTTCCCCGTCGACCGT 14267

Qy        121 GGCTGGGACCTGGACACGCTCTACGACCCCGACCCCGACCACCCCGGCACCTGCTACACC 180
              ||||||||| | ||   ||| ||||||||||| || ||        ||      | |||
Db      14268 GGCTGGGACGTCGAGGGGCTGTACGACCCCGATCCGGAGGCGGTGGGGCGTAGTTACGTG 14327

Qy        181 CGAAACGGCGGATTCCTCTACGACGCAGGCCACTTCGACGCCGAATTCTTCGGCATCAGC 240
              ||  | ||||| ||||| ||   ||  | | | |||||||| || ||||||||  |||
Db      14328 CGGGAGGGCGGGTTCCTGCACTCGGCGGCCGAGTTCGACGCGGAGTTCTTCGGGATCTCG 14387

Qy        241 CCCCGCGAAGCCCTCGCCATGGACCCCCAGCAACGACTCCTCCTCGAAACCGCCTGGGAA 300
              ||||| || ||    || ||||| || ||||| ||  |  || || || ||  | |||||
Db      14388 CCCCGTGAGGCGGCGGCGATGGATCCGCAGCAGCGGTTGCTGCTGGAGACGTCGTGGGAG 14447

Qy        301 ACCATCGAACACGCCGGCATCAACCCCCACACCCTCCACGGCACCCCCACCGGAGTCTTC 360
              |  | || |  || || |||  ||||    | || | |||||||  ||||| |||||
Db      14448 GCGCTGGAGCGGGCGGGGATCGTCCCCGCGTCGCTGCGCGGCACCCGTACCGGCGTCTTC 14507

Qy        361 ACCGGCACCAACGGACAGGACTACGCACTTCGCGTGCACAACGCGGGCCAGTCAACCGAT 420
              ||||||  ||      | |||||||     || |  ||  ||    | |     |||
Db      14508 ACCGGCGTCATGTACGACGACTACGGGTCGCGGTTCGACTCGGCTCCGCCGGAGTACGAG 14567

Qy        421 GGTTTCGCACTGACCGGAACCGCCGGCAGCGTCATCTCCGGTCGTATCTCGTACACGTTT 480
              || | |    ||| ||| | |||||||||| ||   ||||||||  |  | ||  ||||
Db      14568 GGCTACCTCGTGAACGGCAGCGCCGGCAGCATCGCGTCCGGTCGGGTTGCCTATGCGTTG 14627

Qy        481 GGTTTTGAGGGTCCTGCGGTGTCGGTGGACACGGCTTGTTCCTCGTCGTTGGTGGCTTTG 540
              || || |||| || ||| || |||||||||||||| |||| ||||||||||||| |||
Db      14628 GGGTTGGAGGGGCCGGCGCTGACGGTGGACACGGCGTGTTCGTCGTCGTTGGTGGCGTTG 14687

Qy        541 CATCTGGCCTGTCAGGCGTTGCGTGCGGGTGAGTGCTCGATGGCGCTTGCCGGGGGTGTG 600
              ||||||||     ||| |||||||    |||||||||      |||| | |||| || |||
Db      14688 CATCTGGCGGTGCAGTCGTTGCGGCGGGGTGAGTGTGATCTGGCGTTGGCCGGTGGGGTG 14747

Qy        601 ACGGTGATGTCGTCTCCGGGTGCCTTCGTGGAGTTTTCGCGGCAGCGGGGTCTGGCCGCG 660
              ||||||||| || | |||   |  |||||||||| |||||||||||||| |||| |||
Db      14748 ACGGTGATGGCGACGCCGACGGTGCTCGTGGAGTTCTCGCGGCAGCGGGGGCTGGCGGCG 14807

Qy        661 GACGGGCATTGCAAGGCGTTCTCGGCGGCGGCGGACGGGACCGGCTGGGGTGAGGGTGTG 720
              |||||||  |||||||||||| ||| ||  |||||||||| |  ||||  |||||||||
Db      14808 GACGGGCGGTGCAAGGCGTTCGCGGAGGGTGCGGACGGGACGGCGTGGGCCGAGGGTGTG 14867

Qy        721 GGGATGCTGCTGGTGGAGCGGCTCTCCGACGCCCATCGCAACGGTCACCGTGTCCTGGCC 780
              ||  |||||||||||||||||||||||||||||  ||||| || || || || |||||
Db      14868 GGCGTGCTGCTGGTGGAGCGGCTCTCCGACGCCCGCCGCAATGGCCATCGGGTGCTGGCG 14927

Qy        781 GTGGTGCGTGGCAGTGCGGTCAACCAGGACGGTGCGAGCAACGGTCTGACCGCGCCCAAC 840
              |||||||| |||||||||||||| |||||||||||||||||||| ||||| ||||| |
Db      14928 GTGGTGCGGGGCAGTGCGGTCAATCAGGACGGTGCGAGCAACGGGCTGACGGCGCCGAGT 14987

Qy        841 GGGCCGTCCCAGCAGCGTGTCATCCGCCAGGCCCTCGCCAACGCCGGCTTGTCGGCCGGT 900
              ||  ||  | ||||||| || |||||  |||| || ||| |||| ||  || || |||
Db      14988 GGTCCTGCGCAGCAGCGGGTGATCCGTGAGGCGCTGGCCGACGCGGGGCTGACGCCCGCC 15047

Qy        901 GATGTCGACGCGGTGGAGGCCCACGGCACCGGCACCACTTTGGGCGACCCGATCGAGGCC 960
              || |||||||||| |||||  |||||||||||||   |  |||||||||| ||||||||
Db      15048 GACGTCGACGCGGTCGAGGCGCACGGCACCGGCACACCCCTGGGCGACCCCATCGAGGCG 15107

Qy        961 CAGGCCCTCCTCGCGACCTACGGACAGGACCGTGCCGGCGAGGGGCCGCTGTGGCTGGGC 1020
                ||  | ||  || ||||| ||    || ||    ||| | || ||| ||||| ||||
Db      15108 GGTGCGTTGCTGGCCACCTATGGCAGTGAGCGCCAGGGCCAAGGTCCGTTGTGGTTGGGG 15167

Qy       1021 TCGGTCAAGTCCAATGTCGGTCACACACAGGCTGCCGCGGGCGTCGCCGGGGTGATCAAG 1080
              ||| | ||||| ||  |||| ||  | ||||| || ||||| || |  || ||||||||
Db      15168 TCGTTGAAGTCGAACATCGGGCATGCGCAGGCGGCTGCGGGTGTGGGTGGCGTGATCAAG 15227

Qy       1081 ATGGTGATGGCGCTGCGGCATGGTCTGCTGCCGCGGACGTTGCATGTGGATGAGCCGTCG 1140
              |||||  |||| ||||||||||   | |||||||||| |||||||||||| |||||||
Db      15228 GTGGTGCAGGCGATGCGGCATGGGTCGTTGCCGCGGACGCTGCATGTGGATGCGCCGTCG 15287

Qy       1141 CCGCATGTGGACTGGTCCGCGGGTGCGGTGCAGCTGCTGACGGAGACGGTGCCCTGGCCC 1200
              || | ||||| ||| |  |||||||||| |||||||||| |||||   | | |||||
Db      15288 TCGAAGGTGGAGTGGGCTTCGGGTGCGGTGGAGCTGCTGACCGAGACCCGGTCGTGGCCG 15347

Qy       1201 GGCGGGGAGGGGCGGCTACGGCGGGCAGGAGTGTCATCATTCGGCGTCAGCGGCACCAAC 1260
```

```
                |  ||| || |||| | |||||||| |  |||||  | ||||| || ||||| ||||||
Db      15348 CGGCGGGTGGAGCGGGTGCGGCGGGCCGCGGTGTCGGCGTTCGGGGTGAGCGGGACCAAC 15407

Qy       1261 GCCCACGTCATCCTCGAAGAAGCACCCGC 1289
              |||||  ||  |||| || ||||| || ||
Db      15408 GCCCATGTGGTCCTGGAGGAAGCGCCGGC 15436


RESULT 8
AAT80414
ID   AAT80414 standard; DNA; 44377 BP.
XX
AC   AAT80414;
XX
DT   27-FEB-1998  (first entry)
XX
DE   Platenolide synthase gene cluster.
XX
KW   Tylactone synthase gene cluster; tylG gene; multifunctional protein;
KW   platenolide synthase gene cluster; platenolide production; srmG gene;
KW   polyketide; tylactone synthesis; antibiotic; tylosin; ss.
XX
OS   Streptomyces ambofaciens.
XX
FH   Key             Location/Qualifiers
FT   CDS             350..14002
FT                   /*tag= a
FT                   /transl_except= (pos:350..352, aa:Met)
FT                   /note= "ORF1 encodes protein shown in AAW22606"
FT   CDS             14046..20036
FT                   /*tag= b
FT                   /note= "ORF2 encodes protein shown in AAW22607"
FT   CDS             20110..31284
FT                   /*tag= c
FT                   /transl_except= (pos:20111..20113, aa:Met)
FT                   /note= "ORF3 encodes protein shown in AAW22608"
FT   CDS             31329..36071
FT                   /*tag= d
FT                   /note= "ORF4 encodes protein shown in AAW22609"
FT   CDS             36155..41830
FT                   /*tag= e
FT                   /note= "ORF5 encodes protein shown in AAW22610"
XX
PN   EP791655-A2.
XX
PD   27-AUG-1997.
XX
PF   19-FEB-1997;   97EP-0301056.
XX
PR   22-FEB-1996;   96US-0012078.
XX
PA   (ELIL ) LILLY & CO ELI.
XX
PI   Dehoff BS,  Kuhstoss SA,  Rosteck PR,  Sutton KL;
XX
DR   WPI; 1997-418046/39.
DR   P-PSDB; AAW22606-W22610.
XX
PT   DNA encoding Streptomyces fradiae tylactone synthase domain - for
PT   production of tylosin-related polyketide compounds
XX
PS   Example 2; Pages 110-134; 220pp; English.
XX
CC   This sequence represents the platenolide synthase gene cluster of the
CC   invention. This sequence is referred to as the srmG gene, and was
CC   isolated from Streptomyces ambofaciens. This sequence encodes the
CC   multi-functional proteins which direct the synthesis of the polyketide
CC   platenolide. Platenolide is the basic building block of the macrolide
CC   antibiotic spiramycin. This sequence was used along with the tylG gene
CC   (see AAT80413) to create a hybrid ORF1 sequence (see AAT80415). The tylG
CC   gene is the tylactone synthase gene cluster of the invention. The tylG
CC   sequence was isolated from Streptomyces fradiae, and encodes
CC   multifunctional proteins which direct the synthesis of the polyketide
CC   tylactone. Tylactone is the basic building block of the antibiotic
CC   tylosin. The hybrid sequence can be used to transform S. ambofaciens
CC   lacking the srmG ORF1 sequence, or S. fradiae lacking the tylG ORF1
CC   sequence, so that they can produce polyketides. The DNA sequence can be
CC   modified so as to alter the type of carboxylic acids incorporated, the
CC   number of carboxylic acids incorporated and/or the post-condensation
CC   reactions performed, thereby resulting in novel tylosin-related
CC   polyketides.
XX
SQ   Sequence 44377 BP; 4965 A; 15552 C; 17381 G; 6479 T; 0 other;

  Query Match           56.8%;  Score 732.2;  DB 18;  Length 44377;
  Best Local Similarity  73.0%;  Pred. No. 1e-113;
  Matches  941;  Conservative    0;  Mismatches  348;  Indels    0;  Gaps    0;

Qy          1 GAGCCGATCGCCATCGTTGGCATGGCCTGTCGTTTCCCCGGCGGAGTGACCTCGGCGGAC 60
              |||||||||| ||||| |||||||| || |||| ||| ||||| ||| | ||| |||||
Db      14148 GAGCCGATCGCGATCGTGGGCATGGCGTGCCGTTACCCGGGCGGTGTGGCGTCGCCGGAC 14207

Qy         61 GACTTCTGGGATCTGATCTCCTCCGAGCAGGACGCGATCGGCGGATTCCCCACCGACCGC 120
              ||| | ||||| ||| |  |  |||     ||||||| ||  ||  ||||||  ||||||
Db      14208 GACCTGTGGGACCTGGTGGCAGCCGGTACGGACGCGGTCTCCGCGTTCCCCGTCGACCGT 14267

Qy        121 GGCTGGGACCTGGACACGCTCTACGACCCCGACCCCGACCACCCCGGCACCTGCTACACC 180
              |||||||| | ||   ||| |||||||||| || ||        ||      | |||
Db      14268 GGCTGGGACGTCGAGGGGCTGTACGACCCCGATCCGGAGGCGGTGGGGCGTAGTTACGTG 14327

Qy        181 CGAAACGGCGGATTCCTCTACGACGCAGGCCACTTCGACGCCGAATTCTTCGGCATCAGC 240
              ||  | ||||| |||||  ||   || | | | |||||||| || |||||||| |||
Db      14328 CGGGAGGGCGGGTTCCTGCACTCGGCGGCCGAGTTCGACGCGGAGTTCTTCGGGATCTCG 14387

Qy        241 CCCCGCGAAGCCCTCGCCATGGACCCCCAGCAACGACTCCTCCTCGAAACCGCCTGGGAA 300
              ||||| || ||    || |||||| || ||||| ||  | || || || ||  | ||||||
Db      14388 CCCCGTGAGGCGGCGGCGATGGATCCGCAGCAGCGGTTGCTGCTGGAGACGTCGTGGGAG 14447

Qy        301 ACCATCGAACACGCCGGCATCAACCCCCACACCCTCCACGGCACCCCCACCGGAGTCTTC 360
              |  | || |  || || |||  ||||    | || | ||||||||  ||||| ||||||
Db      14448 GCGCTGGAGCGGGCGGGGATCGTCCCCGCGTCGCTGCGCGGCACCCGTACCGGCGTCTTC 14507

Qy        361 ACCGGCACCAACGGACAGGACTACGCACTTCGCGTGCACAACGCGGGCCAGTCAACCGAT 420
              ||||||  ||      | |||||||     || |  ||  ||   | |    |||
Db      14508 ACCGGCGTCATGTACGACGACTACGGGTCGCGGTTCGACTCGGCTCCGCCGGAGTACGAG 14567

Qy        421 GGTTTCGCACTGACCGGAACCGCCGGCAGCGTCATCTCCGGTCGTATCTCGTACACGTTT 480
              || | |   ||| ||| | |||||||||| ||   |||||||  |  | ||  ||||
Db      14568 GGCTACCTCGTGAACGGCAGCGCCGGCAGCATCGCGTCCGGTCGGGTTGCCTATGCGTTG 14627

Qy        481 GGTTTTGAGGGTCCTGCGGTGTCGGTGGACACGGCTTGTTCCTCGTCGTTGGTGGCTTTG 540
              || || |||||  || ||| || |||||||||||| ||||| |||||||||||||| |||
Db      14628 GGGTTGGAGGGGCCGGCGCTGACGGTGGACACGGCGTGTTCGTCGTCGTTGGTGGCGTTG 14687

Qy        541 CATCTGGCCTGTCAGGCGTTGCGTGCGGGTGAGTGCTCGATGGCGCTTGCCGGGGGTGTG 600
              |||||||    ||| |||||||  |||||||     ||||| | ||||| || |||
Db      14688 CATCTGGCGGTGCAGTCGTTGCGGCGGGGTGAGTGTGATCTGGCGTTGGCCGGTGGGGTG 14747

Qy        601 ACGGTGATGTCGTCTCCGGGTGCCTTCGTGGAGTTTTCGCGGCAGCGGGGTCTGGCCGCG 660
              ||||||||  || | |||    | ||||||||| ||||||||||||| ||||| |||
Db      14748 ACGGTGATGGCGACGCCGACGGTGCTCGTGGAGTTCTCGCGGCAGCGGGGGCTGGCGGCG 14807

Qy        661 GACGGGCATTGCAAGGCGTTCTCGGCGGCGGCGGACGGGACCGGCTGGGGTGAGGGTGTG 720
              |||||||  |||||||||||| ||| ||  |||||||||| |   ||||  |||||||||
Db      14808 GACGGGCGGTGCAAGGCGTTCGCGGAGGGTGCGGACGGGACGGCGTGGGCCGAGGGTGTG 14867

Qy        721 GGGATGCTGCTGGTGGAGCGGCTCTCCGACGCCCATCGCAACGGTCACCGTGTCCTGGCC 780
              ||  |||||||||||||||||||||||||||||||||  ||||| || || || || |||||
Db      14868 GGCGTGCTGCTGGTGGAGCGGCTCTCCGACGCCCGCCGCAATGGCCATCGGGTGCTGGCG 14927

Qy        781 GTGGTGCGTGGCAGTGCGGTCAACCAGGACGGTGCGAGCAACGGTCTGACCGCGCCCAAC 840
              |||||||| |||||||||||||| |||||||||||||||||||| ||||| ||||| |
Db      14928 GTGGTGCGGGGCAGTGCGGTCAATCAGGACGGTGCGAGCAACGGGCTGACGGCGCCGAGT 14987

Qy        841 GGGCCGTCCCAGCAGCGTGTCATCCGCCAGGCCCTCGCCAACGCCGGCTTGTCGGCCGGT 900
              || ||  | |||||||| || |||||  |||| || ||| |||| ||  || || |||
```

```
Db      14988 GGTCCTGCGCAGCAGCGGGTGATCCGTGAGGCGCTGGCCGACGCGGGGCTGACGCCCGCC 15047

Qy        901 GATGTCGACGCGGTGGAGGCCCACGGCACCGGCACCACTTTGGGCGACCCGATCGAGGCC 960

Db      15048 GACGTCGACGCGGTCGAGGCGCACGGCACCGGCACACCCCTGGGCGACCCCATCGAGGCG 15107

Qy        961 CAGGCCCTCCTCGCGACCTACGGACAGGACCGTGCCGGCGAGGGGCCGCTGTGGCTGGGC 1020

Db      15108 GGTGCGTTGCTGGCCACCTATGGCAGTGAGCGCCAGGGCCAAGGTCCGTTGTGGTTGGGG 15167

Qy       1021 TCGGTCAAGTCCAATGTCGGTCACACACAGGCTGCCGCGGGCGTCGCCGGGGTGATCAAG 1080

Db      15168 TCGTTGAAGTCGAACATCGGGCATGCGCAGGCGGCTGCGGGTGTGGGTGGCGTGATCAAG 15227

Qy       1081 ATGGTGATGGCGCTGCGGCATGGTCTGCTGCCGCGGACGTTGCATGTGGATGAGCCGTCG 1140

Db      15228 GTGGTGCAGGCGATGCGGCATGGGTCGTTGCCGCGGACGCTGCATGTGGATGCGCCGTCG 15287

Qy       1141 CCGCATGTGGACTGGTCCGCGGGTGCGGTGCAGCTGCTGACGGAGACGGTGCCCTGGCCC 1200

Db      15288 TCGAAGGTGGAGTGGGCTTCGGGTGCGGTGGAGCTGCTGACCGAGACCCGGTCGTGGCCG 15347

Qy       1201 GGCGGGGAGGGGCGGCTACGGCGGGCAGGAGTGTCATCATTCGGCGTCAGCGGCACCAAC 1260

Db      15348 CGGCGGGTGGAGCGGGTGCGGCGGGCCGCGGTGTCGGCGTTCGGGGTGAGCGGGACCAAC 15407

Qy       1261 GCCCACGTCATCCTCGAAGAAGCACCCGC 1289

Db      15408 GCCCATGTGGTCCTGGAGGAAGCGCCGGC 15436


RESULT 9
AAT80415
ID   AAT80415 standard; DNA; 13987 BP.
XX
AC   AAT80415;
XX
DT   02-MAR-1998  (first entry)
XX
DE   Hybrid srmG/tylG ORF1.
XX
KW   Tylactone synthase gene cluster; tylG gene; multifunctional protein;
KW   platenolide synthase gene cluster; platenolide production; srmG gene;
KW   polyketide; tylactone synthesis; antibiotic; tylosin; hybrid gene; ss.
XX
OS   Streptomyces ambofaciens.
OS   Streptomyces fradiae.
XX
FH   Key             Location/Qualifiers
FT   CDS             350..13987
FT                   /*tag= a
FT                   /transl_except= (pos:350..352, aa:Met)
FT                   /note= "ORF1 encodes hybrid protein shown in AAW22611"
XX
PN   EP791655-A2.
XX
PD   27-AUG-1997.
XX
PF   19-FEB-1997;   97EP-0301056.
XX
PR   22-FEB-1996;   96US-0012078.
XX
PA   (ELIL ) LILLY & CO ELI.
XX
PI   Dehoff BS,  Kuhstoss SA,  Rosteck PR,  Sutton KL;
XX
DR   WPI; 1997-418046/39.
DR   P-PSDB; AAW22611.
XX
PT   DNA encoding Streptomyces fradiae tylactone synthase domain - for
PT   production of tylosin-related polyketide compounds
XX
PS   Claim 22; Pages 178-197; 220pp; English.

XX
CC   This sequence represents a hybrid gene of the invention. This sequence
CC   was created by replacing a EcoRI-ApaI fragment of srmG ORF1 with a
CC   EcoRI-SfuI fragment from tylG ORF1. The position of the nucleotides from
CC   each of the two genes is not given in the specification. The srmG gene
CC   (see AAT80414) was isolated from Streptomyces ambofaciens, and encodes
CC   the multi-functional proteins which direct the synthesis of the
CC   polyketide platenolide. Platenolide is the basic building block of the
CC   macrolide antibiotic spiramycin. The tylG gene (see AAT80413) is the
CC   tylactone synthase gene cluster of the invention. The tylG sequence was
CC   isolated from Streptomyces fradiae, and encodes multifunctional proteins
CC   which direct the synthesis of the polyketide tylactone. Tylactone is the
CC   basic building block of the antibiotic tylosin. The hybrid sequence can
CC   be used to transform S. ambofaciens lacking the srmG ORF1 sequence, or S.
CC   fradiae lacking the tylG ORF1 sequence, so that they can produce
CC   polyketides. The DNA sequence can be modified so as to alter the type of
CC   carboxylic acids incorporated, the number of carboxylic acids
CC   incorporated and/or the post-condensation reactions performed, thereby
CC   resulting in novel tylosin-related polyketides.
XX
SQ   Sequence 13987 BP; 1556 A; 4401 C; 5727 G; 2303 T; 0 other;

  Query Match            56.0%;  Score 722.6;  DB 18;  Length 13987;
  Best Local Similarity  73.1%;  Pred. No. 4.4e-112;
  Matches  942;  Conservative    0;  Mismatches  344;  Indels    3;  Gaps    1;

Qy          1 GAGCCGATCGCCATCGTTGGCATGGCCTGTCGTTTCCCCGGCGGAGTGACCTCGGCGGAC 60

Db       8255 GAGCCGATCGCGATCGTGGGCATGGCGTGCCGTTACCCGGGCGGTGTGGCGTCGCCGGAC 8314

Qy         61 GACTTCTGGGATCTGATCTCCTCCGAGCAGGACGCGATCGGCGGATTCCCCACCGACCGC 120

Db       8315 GACCTGTGGGACCTGGTGGCCGGTGACGGGCACACGCTCTCCCCGTTCCCGGCCGACCGT 8374

Qy        121 GGCTGGGACCTGGACACGCTCTACGACCCCGACCCCGACCACCCCGGCACCTGCTACACC 180

Db       8375 GGCTGGGACGTCGAGGGGCTGTACGACCCGGAGCCGGGGGTGCCGGGCAAGAGCTATGTA 8434

Qy        181 CGAAACGGCGGATTCCTCTACGACGCAGGCCACTTCGACGCCGAATTCTTCGGCATCAGC 240

Db       8435 CGGGAAGGCGGGTTCCTGCGTTCCGCGGCCGAGTTCGACGCGGAGTTCTTCGGGATATCG 8494

Qy        241 CCCCGCGAAGCCCTCGCCATGGACCCCCAGCAACGACTCCTCCTCGAAACCGCCTGGGAA 300

Db       8495 CCGCGCGAGGCCACGGCCATGGACCCGCAGCAGCGGTTGCTGCTGGAGACGTCGTGGGAG 8554

Qy        301 ACCATCGAACACGCCGGCATCAACCCCCACACCCTCCACGGCACCCCCACCGGAGTCTTC 360

Db       8555 GCGCTGGAGCGGGCCGGCATCGTTCCGGACTCGCTGCGCGGCACCCGGACCGGTGTCTTC 8614

Qy        361 ACCGGCACCAACGGACAGGACTACGCACTTCGCGTGCACAACGCGGGCCAGTCAACCGAT 420

Db       8615 AGCGGCATCTCCCAGCAGGACTACGCGACCCAGCTGGGGGACGCCGCCGACACCTACGGC 8674

Qy        421 GGTTTCGCACTGACCGGAACCGCCGGCAGCGTCATCTCCGGTCGTATCTCGTACACGTTT 480

Db       8675 GGGCATGTGCTCACGGGGACCCTCGGCAGTGTGATCTCCGGTCGGGTTGCCTATGCGTTG 8734

Qy        481 GGTTTTGAGGGTCCTGCGGTGTCGGTGGACACGGCTTGTTCCTCGTCGTTGGTGGCTTTG 540

Db       8735 GGGTTGGAGGGGCCGGCGCTGACGGTGGACACGGCGTGTTCGTCGTCGTTGGTGGCGTTG 8794

Qy        541 CATCTGGCCTGTCAGGCGTTGCGTGCGGGTGAGTGCTCGATGGCGCTTGCCGGGGGTGTG 600

Db       8795 CATCTGGCGGTGCAGTCGTTGCGGCGGGGTGAGTGTGATCTGGCGTTGGCCGGTGGGGTG 8854

Qy        601 ACGGTGATGTCGTCTCCGGGTGCCTTCGTGGAGTTTTCGCGGCAGCGGGGTCTGGCCGCG 660

Db       8855 ACGGTGATGGCGACGCCGACGGTGTTCGTGGAGTTCTCGCGGCAGCGGGGGCTGGCGGCG 8914

Qy        661 GACGGGCATTGCAAGGCGTTCTCGGCGGCGGCGGACGGGACCGGCTGGGGTGAGGGTGTG 720

Db       8915 GACGGGCGGTGCAAGGCGTTCGCGGAGGGTGCGGACGGGACGGCGTGGGCGGAGGGTGTG 8974
```

```
Qy       721 GGGATGCTGCTGGTGGAGCGGCTCTCCGACGCCCATCGCAACGGTCACCGTGTCCTGGCC 780
            ||  |||||||||||||||||||| |||||||| |  |||||||||| || || ||||||
Db      8975 GGTGTGCTGCTGGTGGAGCGGCTTTCCGACGCGCGCCGCAACGGTCATCGGGTGCTGGCG 9034

Qy       781 GTGGTGCGTGGCAGTGCGGTCAACCAGGACGGTGCGAGCAACGGTCTGACCGCGCCCAAC 840
            |||||||| |||||||||||||| |||||||||||||||||| || ||||| ||||| |
Db      9035 GTGGTGCGGGGCAGTGCGGTCAATCAGGACGGTGCGAGCAATGGGCTGACGGCGCCGAGT 9094

Qy       841 GGGCCGTCCCAGCAGCGTGTCATCCGCCAGGCCCTCGCCAACGCCGGCTTGTCGGCCGGT 900
            || ||| | |||||||| || |||||  |||| || ||  | || ||  ||  | |||
Db      9095 GGTCCGGCGCAGCAGCGGGTGATCCGTGAGGCGCTGGCTGATGCGGGGCTGGTGCCCGCC 9154

Qy       901 GATGTCGACGCGGTGGAGGCCCACGGCACCGGCACCACTTTGGGCGACCCGATCGAGGCC 960
            || || || | |||||||||| ||||| || || ||  |  |||| || ||||||||||
Db      9155 GACGTGGATGTGGTGGAGGCGCACGGTACGGGGACGGCGCTGGGTGATCCGATCGAGGCG 9214

Qy       961 CAGGCCCTCCTCGCGACCTACGGACAGGACCGTGCCGGCGAGGGGCCGCTGTGGCTGGGC 1020
               || || || || || ||||| | ||| || | ||||||    ||| ||||||| ||
Db      9215 GGTGCGCTGCTGGCCACGTACGGGCGGGAGCGGGTCGGCGA---TCCGTTGTGGCTCGGG 9271

Qy      1021 TCGGTCAAGTCCAATGTCGGTCACACACAGGCTGCCGCGGGCGTCGCCGGGGTGATCAAG 1080
            ||| | ||||| ||  |||| ||  | ||||| || ||||| || |  || |||||||||
Db      9272 TCGTTGAAGTCGAACATCGGGCATGCGCAGGCGGCTGCGGGTGTGGGTGGTGTGATCAAG 9331

Qy      1081 ATGGTGATGGCGCTGCGGCATGGTCTGCTGCCGCGGACGTTGCATGTGGATGAGCCGTCG 1140
              |||||  || | ||||||||||    | |||||||||| |||||||||||| |||||||
Db      9332 GTGGTGCAGGGGATGCGGCATGGGTCGTTGCCGCGGACGCTGCATGTGGATGCGCCGTCG 9391

Qy      1141 CCGCATGTGGACTGGTCCGCGGGTGCGGTGCAGCTGCTGACGGAGACGGTGCCCTGGCCC 1200
              || | ||||| ||| |  |||||||||| |||||||||| |||||    | | |||||
Db      9392 TCGAAGGTGGAGTGGGCTTCGGGTGCGGTGGAGCTGCTGACCGAGACCCGGTCGTGGCCG 9451

Qy      1201 GGCGGGGAGGGGCGGCTACGGCGGGCAGGAGTGTCATCATTCGGCGTCAGCGGCACCAAC 1260
              |  ||| || |||| | |||||||| |  |||||  | ||||| || ||||| ||||||
Db      9452 CGGCGGGTGGAGCGGGTGCGGCGGGCCGCGGTGTCGGCGTTCGGGGTGAGCGGGACCAAC 9511

Qy      1261 GCCCACGTCATCCTCGAAGAAGCACCCGC 1289
            ||||| ||  |||| || ||||| || ||
Db      9512 GCCCATGTGGTCCTGGAGGAAGCGCCGGC 9540


RESULT 10
AAD17184
ID   AAD17184 standard; DNA; 65140 BP.
XX
AC   AAD17184;
XX
DT   29-NOV-2001  (first entry)
XX
DE   Streptomyces noursei nys1 DNA of nystatin PKS gene cluster.
XX
KW   Polyketide synthase; PKS; macrolide; nystatin; PKS gene cluster;
KW   antifungal; antibiotic; nys1; ds.
XX
OS   Streptomyces noursei.
XX
FH   Key             Location/Qualifiers
FT   CDS             complement (1..1035)
FT                   /*tag= a
FT                   /product= "NysD2 partial protein"
FT                   /note= "CDS does not include stop codon"
FT   CDS             complement (1056..2576)
FT                   /*tag= b
FT                   /product= "NysD1 protein"
FT   CDS             2806..6906
FT                   /*tag= c
FT                   /product= "NysA protein"
FT   CDS             6952..16530
FT                   /*tag= d
FT                   /product= "NysB protein"
FT   CDS             16550..49840
FT                   /*tag= e
FT                   /product= "NysC protein"
FT   CDS             50260..51015
FT                   /*tag= f
FT                   /product= "NysE protein"
FT   CDS             51405..54305
FT                   /*tag= g
FT                   /product= "NysR1 protein"
FT   CDS             54329..57190
FT                   /*tag= h
FT                   /product= "NysR2 protein"
FT                   /note= "CDS does not include start codon"
FT   CDS             57180..59963
FT                   /*tag= i
FT                   /product= "NysR3 protein"
FT   CDS             60415..61047
FT                   /*tag= j
FT                   /product= "NysR4 (short) protein"
FT                   /note= "CDS does not include start codon"
FT   CDS             61736..62497
FT                   /*tag= k
FT                   /product= "NysR5 protein"
FT                   /note= "CDS does not include start codon"
FT   CDS             complement (62551..63615)
FT                   /*tag= l
FT                   /product= "ORF2 protein"
FT                   /note= "CDS does not include start codon"
FT   CDS             63765..64961
FT                   /*tag= m
FT                   /product= "ORF1 protein"
XX
PN   WO200159126-A2.
XX
PD   16-AUG-2001.
XX
PF   08-FEB-2001; 2001WO-GB00509.
XX
PR   08-FEB-2000; 2000GB-0002840.
PR   10-APR-2000; 2000GB-0008786.
PR   14-APR-2000; 2000GB-0009387.
XX
PA   (UYNO-) UNIV NORGES TEKNISK NATURVITENSKAPELIGE.
PA   (SNTF ) SINTEF STIFTELSEN IND TEK FORSK.
PA   (ALPH-) ALPHARMA AS.
PA   (SINV-) SINVENT AS.
PA   (DZIE/) DZIEGLEWSKA H.
PA   (ZOTC/) ZOTCHEV S B.
PA   (SEKU/) SEKUROVA O N.
PA   (FJAE/) FJAERVIK E.
PA   (BRAU/) BRAUTASET T.
PA   (STRO/) STROM A R.
XX
PI   Zotchev SB,  Sekurova ON,  Fjaervik E,  Brautaset T,  Strom AR;
PI   Valla S,  Ellingsen TE,  Sletta H,  Gulliksen O;
XX
DR   WPI; 2001-557614/62.
DR   P-PSDB; AAE10125, AAE10126, AAE10127, AAE10128, AAE10129, AAE10130,
DR   AAE10131, AAE10132, AAE10133, AAE10134, AAE10135, AAE10136, AAE10137.
XX
PT   New nystatin polyketide synthase polynucleotides and polypeptides,
PT   useful as antibiotics and antifungals -
XX
PS   Claim 2; Page 116-151; 266pp; English.
XX
CC   The present invention relates to the cloning and sequencing of the gene
CC   cluster encoding a modular type I polyketide synthase (PKS) enzyme
CC   involved in the biosynthesis of the macrolide antibiotic nystatin.
CC   The nystatin PKS is useful as antifungal antibiotics. The present
CC   sequence is a Streptomyces noursei nys1 DNA of nystatin PKS gene cluster.
XX
SQ   Sequence 65140 BP; 8270 A; 25171 C; 22273 G; 9426 T; 0 other;

  Query Match             55.5%;  Score 716;  DB 22;  Length 65140;
  Best Local Similarity   73.3%;  Pred. No. 5e-111;
```

```
  Matches  946;  Conservative    0;  Mismatches  335;  Indels    9;  Gaps    2;

Qy          1 GAGCCGATCGCCATCGTTGGCATGGCCTGTCGTTTCCCCGGCGGAGTGACCTCGGCGGAC 60
              || ||||||| |||||| |||||||||||| || | |||||||||  |  ||||  | ||
Db      21818 GACCCGATCGTCATCGTCGGCATGGCCTGCCGCTACCCCGGCGGCATCGCCTCACCCGAG 21877

Qy         61 GACTTCTGGGATCTGATCTCCTCCGAGCAGGACGCGATCGGCGGATTCCCCACCGACCGC 120
              ||| |||||   ||| ||  |   |      |||||| |  |||   |||||||||| |||||
Db      21878 GACCTCTGGCGCCTGGTCAGCCAGGGCGCCGACGCCACTGGCCCGTTCCCCACCAACCGC 21937

Qy        121 GGCTGGGACCTGGACACGCTCTACGACCCCGACCCCGACCACCCCGGCACCTGCTACACC 180
              ||||||||||||||||  |||||||||||||||||||||| ||| |||  |  | ||  |
Db      21938 GGCTGGGACCTGGACAACCTCTACGACCCCGACCCCGACCGCCCGGGCCGCACCCACGTC 21997

Qy        181 CGAAACGGCGGATTCCTCTACGACGCAGGCCACTTCGACGCCGAATTCTTCGGCATCAGC 240
              ||   |||||| |||||  ||||||| |||  |||||||||||| |||||||| || |||  .
Db      21998 CGCGCCGGCGGCTTCCTGCACGACGCCGGCTCCTTCGACGCCGACTTCTTCGGGATGAGC 22057

Qy        241 CCCCGCGAAGCCCTCGCCATGGACCCCCAGCAACGACTCCTCCTCGAAACCGCCTGGGAA 300
              || ||||| ||  | |||| |||| ||||||| || || || ||||||  | ||||||||
Db      22058 CCGCGCGAGGCGATGGCCACCGACTCCCAGCAGCGCCTGCTGCTCGAACTCTCCTGGGAA 22117

Qy        301 ACCATCGAACACGCCGGCATCAACCCCCACACCCTCCACGGCACCCCCACCGGAGTCTTC 360
              || |||||| |||||||||| |||||  | | || | || | ||  ||||| ||||||
Db      22118 GCCGTCGAACGCGCCGGCATCGACCCCGCCTCACTGCGCGACTCCGGCACCGGCGTCTTC 22177

Qy        361 ACCGGCACCAACGGACAGGACTACGCACTTCGCGTGCACAACGCGGGCCAGTCAACCGAT 420
              |||||  ||      | |||||||     ||      |   ||| |     |||
Db      22178 GCCGGCGTCATGTACAACGACTACG------GCACCACCCTGACCGGCGACGAGTACGAG 22231

Qy        421 GGTTTCGCACTGACCGGAACCGCCGGCAGCGTCATCTCCGGTCGTATCTCGTACACGTTT 480
              |  |||      | ||| | ||||   ||||||  ||||| ||  |||| |||||  |
Db      22232 GCGTTCCGCGGCAACGGCAGCGCCCCGAGCGTCGCCTCCGGCCGCGTCTCCTACACCCTC 22291

Qy        481 GGTTTTGAGGGTCCTGCGGTGTCGGTGGACACGGCTTGTTCCTCGTCGTTGGTGGCTTTG 540
              ||  | || || || || ||  |||||||||| || || || || ||  |||| ||  ||
Db      22292 GGCCTGGAAGGCCCGGCCGTCACGGTGGACACCGCCTGCTCTTCCTCTCTGGTCGCCCTG 22351

Qy        541 CATCTGGCCTGTCAGGCGTTGCGTGCGGGTGAGTGCTCGATGGCGCTTGCCGGGGGTGTG 600
              ||   |||    |||||||||| ||||| ||||||||| ||||| | |||||| ||||||
Db      22352 CACTGGGCGGCGCAGGCGTTGCGGGCGGGGGAGTGCTCGTTGGCGTTGGCCGGTGGTGTG 22411

Qy        601 ACGGTGATGTCGTCTCCGGGTGCCTTCGTGGAGTTTTCGCGGCAGCGGGGTCTGGCCGCG 660
              |||||||||||| | ||| | | |||||||||| ||||||||||||||||||||  |
Db      22412 ACGGTGATGTCGACGCCGAGCACGTTCGTGGAGTTCTCGCGGCAGCGGGGTCTGGCGCCT 22471

Qy        661 GACGGGCATTGCAAGGCGTTCTCGGCGGCGGCGGACGGGACCGGCTGGGGTGAGGGTGTG 720
              || || | ||  |||||||| | | ||| |||||||     | ||||   ||||| ||
Db      22472 GATGGTCGTTCGAAGGCGTTCGCCGAGGCCGCGGACGGCGTGGCCTGGTCCGAGGGCGTC 22531

Qy        721 GGGATGCTGCTGGTGGAGCGGCTCTCCGACGCCCATCGCAACGGTCACCGTGTCCTGGCC 780
              || ||||||  |  |||||||||  || |||||    ||||||||||||    ||||||||
Db      22532 GGCATGCTGGTCCTGGAGCGGCAGTCGGACGCGGTGCGCAACGGTCACGAGATCCTGGCC 22591

Qy        781 GTGGTGCGTGGCAGTGCGGTCAACCAGGACGGTGCGAGCAACGGTCTGACCGCGCCCAAC 840
              ||||||||  |||   |||||||||||||||||||||  ||||||||||||||||||||
Db      22592 GTGGTGCGCGGCTCGGCGGTCAACCAGGACGGTGCGTCCAACGGTCTGACCGCGCCCAAC 22651

Qy        841 GGGCCGTCCCAGCAGCGTGTCATCCGCCAGGCCCTCGCCAACGCCGGCTTGTCGGCCGGT 900
              || |||||||||||||| || ||||| |||||  | ||||  | ||||  ||||   | |
Db      22652 GGCCCGTCCCAGCAGCGGGTGATCCGTCAGGCGTTGGCCAGTGGCGGCCTGTCCACGGCC 22711

Qy        901 GATGTCGACGCGGTGGAGGCCCACGGCACCGGCACCACTTTGGGCGACCCGATCGAGGCC 960
              || || ||||| || ||||| |||||||| || || ||  | || ||||||||||||||||
Db      22712 GACGTGGACGCCGTTGAGGCGCACGGCACGGGTACGACGCTCGGTGACCCGATCGAGGCC 22771

Qy        961 CAGGCCCTCCTCGCGACCTACGGACAGGACCGTGCCGGCGAGGGGCCGCTGTGGCTGGGC 1020
              ||||| ||||| || |||||||| |  ||||| | |  ||||   ||||||  ||| |||
Db      22772 CAGGCGCTCCTGGCCACCTACGGTCGCGACCGCGACCCCGAGAACCCGCTGCTGCTCGGC 22831

Qy       1021 TCGGTCAAGTCCAATGTCGGTCACACACAGGCTGCCGCGGGCGTCGCCGGGGTGATCAAG 1080
              ||| ||||||||||  |||||||||| ||||| || || || |||||||| || ||||||
Db      22832 TCGATCAAGTCCAACATCGGTCACACCCAGGCAGCGGCCGGTGTCGCCGGTGTCATCAAG 22891

Qy       1081 ATGGTGATGGCGCTGCGGCATGGTCTGCTGCCGCGGACGTTGCATGTGGATGAGCCGTCG 1140
              ||||| |||||| ||||||| ||  |||||||| |||  ||||||| || | ||||||
Db      22892 ATGGTCATGGCGATGCGGCACGGCGTGCTGCCGCAGACCCTGCATGTCGACGCGCCGTCC 22951

Qy       1141 CCGCATGTGGACTGGTCCGCGGGTGCGGTGCAGCTGCTGACGGAGACGGTGCCCTGGCCC 1200
              |||| || || |||  ||  || || ||  | ||||| || |||  |    |||||||
Db      22952 TCGCACGTCGATTGGAGCGTCGGCGCCGTCGAACTGCTCACCGAGCAGACCGCCTGGCC- 23010

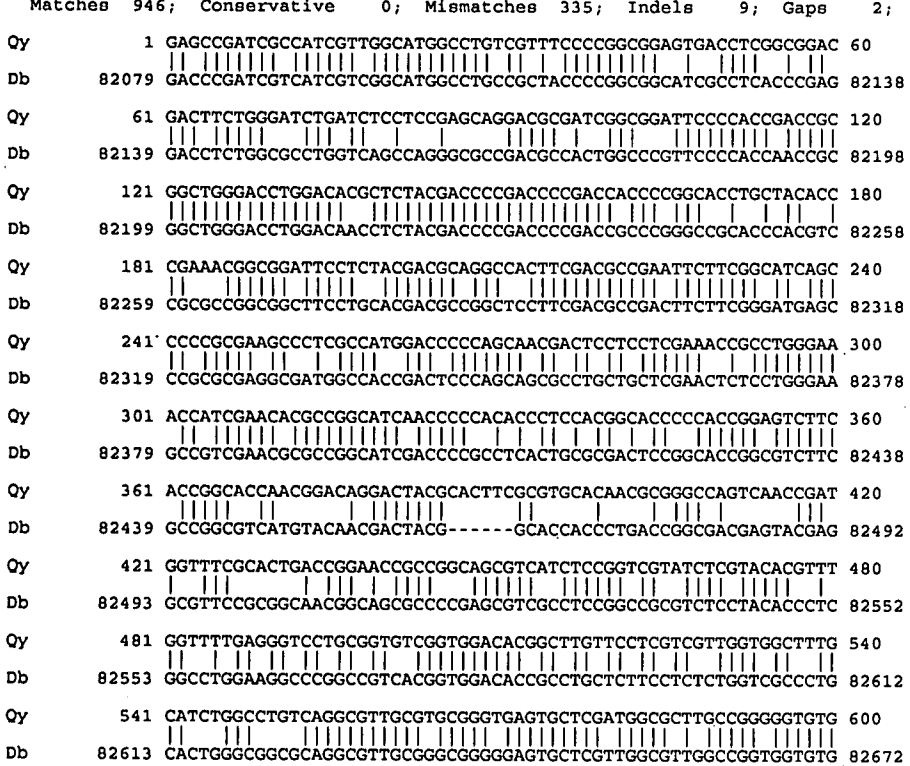
Qy       1201 GGCGGGGAGGGGCGGCTACGGCGGGCAGGAGTGTCATCATTCGGCGTCAGCGGCACCAAC 1260
              | |   || |||   || || || || || || || || |||||| ||||||||||||||
Db      23011 --GGAGACCGGCCGGGCCCGTCGCGCCGGTGTCTCCTCCTTCGGCATCAGCGGCACCAAC 23068

Qy       1261 GCCCACGTCATCCTCGAAGAAGCACCCGCC 1290
              ||||||||| || ||||  |  | ||  ||
Db      23069 GCCCACGTCGTCATCGAGCAGTCCCCGACC 23098


RESULT 11
AAD17186
ID   AAD17186 standard; DNA; 125401 BP.
XX
AC   AAD17186;
XX
DT   29-NOV-2001  (first entry)
XX
DE   Streptomyces noursei nystatin PKS gene cluster DNA.
XX
KW   Polyketide synthase; PKS; macrolide; nystatin; PKS gene cluster;
KW   antifungal; antibiotic; ds.
XX
OS   Streptomyces noursei.
XX
FH   Key             Location/Qualifiers
FT   CDS             6337..34771
FT                   /*tag=  a
FT                   /product= "NysI complete protein"
FT   CDS             34792..51099
FT                   /*tag=  b
FT                   /product= "NysJ protein"
FT   CDS             51155..57355
FT                   /*tag=  c
FT                   /product= "NysK protein"
FT   CDS             57503..58687
FT                   /*tag=  d
FT                   /product= "NysL protein"
FT   CDS             complement (58786..58980)
FT                   /*tag=  e
FT                   /product= "NysM protein"
FT                   /note= "CDS does not include start codon"
FT   CDS             complement (59045..60241)
FT                   /*tag=  f
FT                   /product= "NysN protein"
FT                   /note= "CDS does not include start codon"
FT   CDS             complement (60238..61296)
FT                   /*tag=  g
FT                   /product= "NysD2 complete protein"
FT   CDS             120628..121308
FT                   /*tag=  h
FT                   /product= "NysR4 (long) protein"
XX
PN   WO200159126-A2.
XX
PD   16-AUG-2001.
XX
PF   08-FEB-2001; 2001WO-GB00509.
XX
PR   08-FEB-2000; 2000GB-0002840.
PR   10-APR-2000; 2000GB-0008786.
PR   14-APR-2000; 2000GB-0009387.
XX
PA   (UYNO-) UNIV NORGES TEKNISK NATURVITENSKAPELIGE.
PA   (SNTF ) SINTEF STIFTELSEN IND TEK FORSK.
```

PA (ALPH-) ALPHARMA AS.
PA (SINV-) SINVENT AS.
PA (DZIE/) DZIEGLEWSKA H.
PA (ZOTC/) ZOTCHEV S B.
PA (SEKU/) SEKUROVA O N.
PA (FJAE/) FJAERVIK E.
PA (BRAU/) BRAUTASET T.
PA (STRO/) STROM A R.
XX
PI Zotchev SB, Sekurova ON, Fjaervik E, Brautaset T, Strom AR;
PI Valla S, Ellingsen TE, Sletta H, Gulliksen O;
XX
DR WPI; 2001-557614/62.
DR P-PSDB; AAE10143, AAE10144, AAE10145, AAE10146, AAE10147, AAE10148,
DR AAE10149, AAE10150.
XX
PT New nystatin polyketide synthase polynucleotides and polypeptides,
PT useful as antibiotics and antifungals -
XX
PS Claim 1; Page 188-254; 266pp; English.
XX
CC The present invention relates to the cloning and sequencing of the gene
CC cluster encoding a modular type I polyketide synthase (PKS) enzyme
CC involved in the biosynthesis of the macrolide antibiotic nystatin.
CC The nystatin PKS is useful as antifungal antibiotics. The present
CC sequence is a Streptomyces noursei nystatin PKS gene cluster DNA.
XX
SQ Sequence 125401 BP; 15664 A; 49692 C; 42871 G; 17174 T; 0 other;

Query Match 55.5%; Score 716; DB 22; Length 125401;
Best Local Similarity 73.3%; Pred. No. 4.8e-111;
Matches 946; Conservative 0; Mismatches 335; Indels 9; Gaps 2;



RESULT 12
AAZ87298
ID AAZ87298 standard; DNA; 11220 BP.
XX
AC AAZ87298;
XX
DT 05-JUN-2000 (first entry)
XX
DE S. venezuelae macrolide biosynthetic gene pikAII, SEQ ID NO:32.
XX
KW Desosamine biosynthesis; macrolide; polyketide; methymycin; pikromycin;
KW neomethymycin; narbomycin; polyhydroxyalkanoate monomer synthase;
KW biopolymer; antibiotic; chemotherapeutic; immunosuppressant; asthma;
KW chronic obstructive pulmonary disease; respiratory inflammation;
KW hypercholesterolaemia; crop protection agent; ds.
XX
OS Streptomyces venezuelae ATCC15439.
XX
FH Key Location/Qualifiers
FT CDS 1..111220
FT /*tag= a
FT /product= "PikAII"
XX
PN WO200000620-A2.

```
XX
PD   06-JAN-2000.
XX
PF   25-JUN-1999;   99WO-US14398.
XX
PR   26-JUN-1998;   98US-0105537.
XX
PA   (MINU ) UNIV MINNESOTA.
XX
PI   Sherman DH,  Liu H,  Xue Y,  Zhao L;
XX
DR   WPI; 2000-160679/14.
DR   P-PSDB; AAY77193.
XX
PT   Desosamine and macrolide biosynthetic gene clusters, useful for, e .g.
PT   synthesis of methymycin and pikromycin -
XX
PS   Claim 15; Page 398-403; 438pp; English.
XX
CC   The invention relates to an isolated and purified nucleic acid segment
CC   comprising a desosamine biosynthetic gene cluster, a fragment or its
CC   biologically active variant, where the nucleic acid sequence is not
CC   derived from the eryC gene cluster of Saccharopolyspora erythraea or
CC   Streptomyces antibioticus. The invention also relates to a macrolide
CC   biosynthetic gene cluster, or fragments thereof. The macrolide
CC   biosynthetic gene cluster encodes proteins which synthesise methymycin,
CC   pikromycin, neomethymycin, narbomycin or a combination of these
CC   compounds. Recombinant or augmented cells comprising the desosamine
CC   and/or macrolide biosynthetic gene clusters are useful for the
CC   production of biologically active macrolides. The macrolide biosynthetic
CC   proteins are useful for synthesis of methymycin, pikromycin,
CC   neomethymycin and narbomycin. The alternative termination of polyketide
CC   synthesis may be useful to prepare novel antibiotics and
CC   polyhydroxyalkanoate (PHA) monomers. The compounds produced by the
CC   recombinant host cells are useful as biopolymers, e.g., in packaging or
CC   biomedical applications, to engineer PHA monomer synthases or to prepare
CC   biologically active agents, such as chemotherapeutics,
CC   immunosuppressants, agents to treat asthma, chronic obstructive pulmonary
CC   disease as well as other diseases involving respiratory inflammation,
CC   cholesterol-lowering agents or macrolide-based antibiotics which are
CC   active against a variety of organisms, e.g., bacteria, including
CC   multi-drug resistant pneumococci and other respiratory pathogens, as well
CC   as viral parasitic pathogens, or as crop protection agents (e.g.,
CC   fungicides or insecticides) via expression of polyketides in plants.
CC   Sequences AAZ87295-Z87302 represent macrolide biosynthetic genes from
CC   Streptomyces venezuelae ATCC 15439, which encode proteins
CC   AAY77190-Y77197.
XX
SQ   Sequence 11220 BP; 1369 A; 4423 C; 3966 G; 1462 T; 0 other;

  Query Match             55.5%;  Score 715.4;  DB 21;  Length 11220;
  Best Local Similarity   72.6%;  Pred. No. 7e-111;
  Matches  939;  Conservative    0;  Mismatches  351;  Indels    3;  Gaps    1;

Qy          1 GAGCCGATCGCCATCGTTGGCATGGCCTGTCGTTTCCCCGGCGGAGTGACCTCGGCGGAC 60
              |||||  ||||  |||||  |||||||| || ||   | || |||||  ||  |||||  ||||
Db       4618 GAGCCCGTCGCGATCGTCGGCATGGCGTGCCGCCTGCCGGGCGGTGTCGCCTCGCCGGAG 4677

Qy         61 GACTTCTGGGATCTGATCTCCTCCGAGCAGGACGCGATCGGCGGATTCCCCACCGACCGC 120
              |||  | |||   |||  |  ||  ||   ||||||||||  |||  |||||    ||||||
Db       4678 GACCTGTGGCGGCTGGTGGCCGGCGGCGAGGACGCGATCTCCGGCTTCCCGCAGGACCGC 4737

Qy        121 GGCTGGGACCTGGACACGCTCTACGACCCCGACCCCGACCACCCCGGCACCTGCTACACC 180
              |||||||||  ||||   |||  |||||||| |||||  |||     ||||        |||  |
Db       4738 GGCTGGGACGTGGAGGGGCTGTACGACCCGGACCCGGACGCGTCCGGGCGGACGTACTGC 4797

Qy        181 CGAAACGGCGGATTCCTCTACGACGCAGGCCACTTCGACGCCGAATTCTTCGGCATCAGC 240
              ||   |||  ||  ||||||  ||||  ||  |||  |  ||||||||||  ||||||||  |||
Db       4798 CGTGCCGGTGGCTTCCTCGACGAGGCGGGCGAGTTCGACGCCGACTTCTTCGGGATCTCG 4857

Qy        241 CCCCGCGAAGCCCTCGCCATGGACCCCCAGCAACGACTCCTCCTCGAAACCGCCTGGGAA 300
              ||  |||||  ||||||||||||||||||  |||||  ||  ||||||||  ||  |||  |||||||
Db       4858 CCGCGCGAGGCCCTCGCCATGGACCCGCAGCAGCGGCTCCTCCTGGAGACCTCCTGGGAG 4917

Qy        301 ACCATCGAACACGCCGGCATCAACCCCCACACCCTCCACGGCACCCCCACCGGAGTCTTC 360
              ||  ||||  |||||||  |||  ||||    |  ||||  ||  ||     |   |||  ||  |||
Db       4918 GCCGTCGAGGACGCCGGGATCGACCCGACCTCCCTTCAGGGGCAGCAGGTCGGCGTGTTC 4977

Qy        361 ACCGGCACCAACGGACAGGACTACGCACTTCGCGTGCACAACGCGGGCCAGTCAACCGAT 420
              |  ||||||||||| |   ||||||  |  |   | | ||||  | |  |  ||       ||
Db       4978 GCGGGCACCAACGGCCCCCACTACGAGCCGCTGCTCCGCAACACCGCCGAGGATCTTGAG 5037

Qy        421 GGTTTCGCACTGACCGGAACCGCCGGCAGCGTCATCTCCGGTCGTATCTCGTACACGTTT 480
              ||||  ||    |||  ||  |  |||||  ||||  ||||  ||  ||  |||  ||||||||||  |
Db       5038 GGTTACGTCGGGACGGGCAACGCCGCCAGCATCATGTCGGGCCGTGTCTCGTACACCCTC 5097

Qy        481 GGTTTTGAGGGTCCTGCGGTGTCGGTGGACACGGCTTGTTCCTCGTCGTTGGTGGCTTTG 540
              ||  |  |||||  ||  ||  ||  ||||  |||||  ||  ||  |||||  |||  ||||  ||  ||
Db       5098 GGCCTGGAGGGCCCGGCCGTCACGGTCGACACCGCCTGCTCCTCCTCGCTGGTCGCCCTG 5157

Qy        541 CATCTGGCCTGTCAGGCGTTGCGTGCGGGTGAGTGCTCGATGGCGCTTGCCGGGGGTGTG 600
              ||  ||  |||   |||||  ||||   |||  ||  |||    |||||||  ||  ||  ||||||
Db       5158 CACCTCGCCGTGCAGGCCCTGCGCAAGGGCGAATGCGGACTGGCGCTCGCGGGCGGTGTG 5217

Qy        601 ACGGTGATGTCGTCTCCGGGTGCCTTCGTGGAGTTTTCGCGGCAGCGGGGTCTGGCCGCG 660
              |||||  ||||||  |  ||      |  ||||||||||     ||||||||  ||  ||  ||  |  |
Db       5218 ACGGTCATGTCGACGCCCACGACGTTCGTGGAGTTCAGCCGGCAGCGCGGGCTCGCGGAG 5277

Qy        661 GACGGGCATTGCAAGGCGTTCTCGGCGGCGGCGGACGGGACCGGCTGGGGTGAGGGTGTG 720
              |||||  |  |  |||||||||  |  |||  ||||||||||     ||||   ||  |||||  ||
Db       5278 GACGGCCGGTCGAAGGCGTTCGCCGCGTCGGCGGACGGCTTCGGCCCGGCGGAGGGCGTC 5337

Qy        721 GGGATGCTGCTGGTGGAGCGGCTCTCCGACGCCCATCGCAACGGTCACCGTGTCCTGGCC 780
              ||  |||||  ||  ||  |||||  ||  ||  |||||||   ||||||||  ||||||||  |||||
Db       5338 GGCATGCTCCTCGTCGAGCGCCTGTCGGACGCCCGCCGCAACGGACACCGTGTGCTGGCG 5397

Qy        781 GTGGTGCGTGGCAGTGCGGTCAACCAGGACGGTGCGAGCAACGGTCTGACCGCGCCCAAC 840
              ||  |||||  |||||||||||||||||||||||||  ||||||||||  ||||||||  ||  |||
Db       5398 GTCGTGCGCGGCAGCGCGGTCAACCAGGACGGCGCGAGCAACGGCCTGACCGCCCCGAAC 5457

Qy        841 GGGCCGTCCCAGCAGCGTGTCATCCGCCAGGCCCTCGCCAACGCCGGCTTGTCGGCCGGT 900
              |||||  ||  ||||||||  ||||||||  |  ||  |||||  |||||  |  ||  ||  |||
Db       5458 GGGCCCTCGCAGCAGCGCGTCATCCGGCGCGCGCTCGCGGACGCCCGACTGACGACCGCC 5517

Qy        901 GATGTCGACGCGGTGGAGGCCCACGGCACCGGCACCACTTTGGGCGACCCGATCGAGGCC 960
              ||  ||  ||||   ||  ||||||||||||||  |||||     |  ||||||||||||||||
Db       5518 GACGTGGACGTCGTCGAGGCCCACGGCACGGGCACGCGACTCGGCGACCCGATCGAGGCA 5577

Qy        961 CAGGCCCTCCTCGCGACCTACGGACAGGACCGTGCCGGCGAGGGGCCGCTGTGGCTGGGC 1020
              |||||||||  ||||  ||||||||  ||||   ||  |  |  |||   |||||||  | |||||
Db       5578 CAGGCCCTCATCGCCACCTACGGCCAGGGGCGCGACACCGAACAGCCGCTGCGCCTGGGG 5637

Qy        1021 TCGGTCAAGTCCAATGTCGGTCACACACAGGCTGCCGCGGGCGTCGCCGGGGTGATCAAG 1080
               |||  |  ||||||||   ||||  |||||  |||||  |||||  ||  |||  ||||   |  ||||||
Db       5638 TCGTTGAAGTCCAACATCGGACACACCCAGGCCGCCGCCGGTGTCTCCGGCATCATCAAG 5697

Qy        1081 ATGGTGATGGCGCTGCGGCATGGTCTGCTGCCGCGGACGTTGCATGTGGATGAGCCGTCG 1140
               |||||   ||||  ||||  ||  ||  |  ||||||  ||||  |  ||  |||||   |||||||
Db       5698 ATGGTCCAGGCGATGCGCCACGGCGTCCTGCCGAAGACGCTCCACGTGGACCGGCCGTCG 5757

Qy        1141 CCGCATGTGGACTGGTCCGCGGGTGCGGTGCAGCTGCTGACGGAGACGGTGCCCTGGCCC 1200
                  ||  |  ||||||||  |||||  ||||  ||||||||  ||  |||  |  ||  ||||||
Db       5758 GACCAGATCGACTGGTCGGCGGGCACGGTCGAGCTGCTCACCGAGGCCATGGACTGGCCG 5817

Qy        1201 GG---CGGGGAGGGGCGGCTACGGCGGGCAGGAGTGTCATCATTCGGCGTCAGCGGCACC 1257
               |     ||||||  ||||  ||  ||  |  |  ||  ||  ||  ||||||  ||||||||||
Db       5818 AGGAAGCAGGAGGGCGGGCTGCGCCGCGCGGCCGTCTCCTCCTTCGGCATCAGCGGCACG 5877

Qy        1258 AACGCCCACGTCATCCTCGAAGAAGCACCCGCC 1290
               |||||  |||  ||  |  ||||||||||||  ||  |  |
Db       5878 AACGCGCACATCGTGCTCGAAGAAGCCCCGGTC 5910


RESULT 13
AAZ87318
ID   AAZ87318 standard; DNA; 36778 BP.
```

```
XX
AC   AAZ87318;
XX
DT   05-JUN-2000  (first entry)
XX
DE   S. venezuelae pik (macrolide biosynthesis) gene cluster.
XX
KW   Desosamine biosynthesis; macrolide; polyketide; methymycin; pikromycin;
KW   neomethymycin; narbomycin; polyhydroxyalkanoate monomer synthase;
KW   biopolymer; antibiotic; chemotherapeutic; immunosuppressant; asthma;
KW   chronic obstructive pulmonary disease; respiratory inflammation;
KW   hypercholesterolaemia; crop protection agent; ds.
XX
OS   Streptomyces venezuelae ATCC15439.
XX
FH   Key             Location/Qualifiers
FT   CDS             1742..15583
FT                   /*tag= a
FT                   /product= "Pik gene cluster protein #1 (AAY77200)"
FT   CDS             15688..26907
FT                   /*tag= b
FT                   /product= "Pik gene cluster protein #2 (AAY77201)"
FT   CDS             26991..31679
FT                   /*tag= c
FT                   /product= "Pik gene cluster protein #3 (AAY77202)"
FT   CDS             31782..35822
FT                   /*tag= d
FT                   /product= "Pik gene cluster protein #4 (AAY77203)"
FT   CDS             35819..36664
FT                   /*tag= e
FT                   /product= "Pik gene cluster protein #4 (AAY80997)"
XX
PN   WO200000620-A2.
XX
PD   06-JAN-2000.
XX
PF   25-JUN-1999;   99WO-US14398.
XX
PR   26-JUN-1998;   98US-0105537.
XX
PA   (MINU ) UNIV MINNESOTA.
XX
PI   Sherman DH,  Liu H,  Xue Y,  Zhao L;
XX
DR   WPI; 2000-160679/14.
DR   P-PSDB; AAY77200, AAY77201, AAY77202, AAY77203, AAY80997.
XX
PT   Desosamine and macrolide biosynthetic gene clusters, useful for, e .g.
PT   synthesis of methymycin and pikromycin -
XX
PS   Disclosure; Figure 31; 438pp; English.
XX
CC   The invention relates to an isolated and purified nucleic acid segment
CC   comprising a desosamine biosynthetic gene cluster, a fragment or its
CC   biologically active variant, where the nucleic acid sequence is not
CC   derived from the eryC gene cluster of Saccharopolyspora erythraea or
CC   Streptomyces antibioticus. The invention also relates to a macrolide
CC   biosynthetic gene cluster, or fragments thereof. The macrolide
CC   biosynthetic gene cluster encodes proteins which synthesise methymycin,
CC   pikromycin, neomethymycin, narbomycin or a combination of these
CC   compounds. Recombinant or augmented cells comprising the desosamine
CC   and/or macrolide biosynthetic gene clusters are useful for the production
CC   of biologically active macrolides. The macrolide biosynthetic proteins
CC   are useful for synthesis of methymycin, pikromycin, neomethymycin and
CC   narbomycin. The alternative termination of polyketide synthesis may be
CC   useful to prepare novel antibiotics and polyhydroxyalkanoate (PHA)
CC   monomers. The compounds produced by the recombinant host cells are useful
CC   as biopolymers, e.g., in packaging or biomedical applications, to
CC   engineer PHA monomer synthases or to prepare biologically active agents,
CC   such as chemotherapeutics, immunosuppressants, agents to treat asthma,
CC   chronic obstructive pulmonary disease as well as other diseases involving
CC   respiratory inflammation, cholesterol-lowering agents or macrolide-based
CC   antibiotics which are active against a variety of organisms, e.g.,
CC   bacteria, including multi-drug resistant pneumococci and other
CC   respiratory pathogens, as well as viral parasitic pathogens, or as crop
CC   protection agents (e.g., fungicides or insecticides) via expression of
CC   polyketides in plants. The present sequence represents the macrolide
CC   biosynthetic gene cluster (pik) from Streptomyces venezuelae ATCC
CC   15439, as given in figure 31.
XX
SQ   Sequence 36778 BP; 4758 A; 14303 C; 12864 G; 4851 T; 2 other;

  Query Match            55.5%;  Score 715.4;  DB 21;  Length 36778;
  Best Local Similarity  72.6%;  Pred. No. 6.5e-111;
  Matches  939;  Conservative    0;  Mismatches  351;  Indels    3;  Gaps    1;

Qy          1 GAGCCGATCGCCATCGTTGGCATGGCCTGTCGTTTCCCCGGCGGAGTGACCTCGGCGGAC 60
              ||||| |||| ||||| |||||||| || ||  | || ||||| ||  |||| |||
Db      20305 GAGCCCGTCGCGATCGTCGGCATGGCGTGCCGCCTGCCGGGCGGTGTCGCCTCGCCGGAG 20364

Qy         61 GACTTCTGGGATCTGATCTCCTCCGAGCAGGACGCGATCGGCGGATTCCCCACCGACCGC 120
              ||| | |||   ||| | ||  ||   |||||||||| ||| |||||    ||||||
Db      20365 GACCTGTGGCGGCTGGTGGCCGGCGGCGAGGACGCGATCTCCGGCTTCCCGCAGGACCGC 20424

Qy        121 GGCTGGGACCTGGACACGCTCTACGACCCCGACCCCGACCACCCCGGCACCTGCTACACC 180
              |||||||| ||||   ||| |||||||| ||||| |||    ||||         |||  |
Db      20425 GGCTGGGACGTGGAGGGGCTGTACGACCCGGACCCGGACGCGTCCGGGCGGACGTACTGC 20484

Qy        181 CGAAACGGCGGATTCCTCTACGACGCAGGCCACTTCGACGCCGAATTCTTCGGCATCAGC 240
              ||   ||| || |||||| |||| || ||| | |||||||||||| |||||||| |||
Db      20485 CGTGCCGGTGGCTTCCTCGACGAGGCGGGCGAGTTCGACGCCGACTTCTTCGGGATCTCG 20544

Qy        241 CCCCGCGAAGCCCTCGCCATGGACCCCCAGCAACGACTCCTCCTCGAAACCGCCTGGGAA 300
              || ||||| |||||||||||||||| ||||| || |||||||| || ||| |||||||
Db      20545 CCGCGCGAGGCCCTCGCCATGGACCCGCAGCAGCGGCTCCTCCTGGAGACCTCCTGGGAG 20604

Qy        301 ACCATCGAACACGCCGGCATCAACCCCCACACCCTCCACGGCACCCCCACCGGAGTCTTC 360
              || ||||  ||||||| ||| ||||   | |||| || ||    |   ||| || |||
Db      20605 GCCGTCGAGGACGCCGGGATCGACCCGACCTCCCTTCAGGGGCAGCAGGTCGGCGTGTTC 20664

Qy        361 ACCGGCACCAACGGACAGGACTACGCACTTCGCGTGCACAACGCGGGCCAGTCAACCGAT 420
              | |||||||||||  |   ||||||  |  |   | | ||||  | | | ||      ||
Db      20665 GCGGGCACCAACGGCCCCCACTACGAGCCGCTGCTCCGCAACACCGCCGAGGATCTTGAG 20724

Qy        421 GGTTTCGCACTGACCGGAACCGCCGGCAGCGTCATCTCCGGTCGTATCTCGTACACGTTT 480
              |||| ||    ||| || | ||||| |||| |||| || || ||| ||||||||||  |
Db      20725 GGTTACGTCGGGACGGGCAACGCCGCCAGCATCATGTCGGGCCGTGTCTCGTACACCCTC 20784

Qy        481 GGTTTTGAGGGTCCTGCGGTGTCGGTGGACACGGCTTGTTCCTCGTCGTTGGTGGCTTTG 540
              ||  | ||||| || || ||  |||| ||||| || || ||||| ||| |||| ||  ||
Db      20785 GGCCTGGAGGGCCCGGCCGTCACGGTCGACACCGCCTGCTCCTCCTCGCTGGTCGCCCTG 20844

Qy        541 CATCTGGCCTGTCAGGCGTTGCGTGCGGGTGAGTGCTCGATGGCGCTTGCCGGGGGTGTG 600
              || || |||   |||||  ||||   ||| || |||    ||||||| || || ||||||
Db      20845 CACCTCGCCGTGCAGGCCCTGCGCAAGGGCGAATGCGGACTGGCGCTCGCGGGCGGTGTG 20904

Qy        601 ACGGTGATGTCGTCTCCGGGTGCCTTCGTGGAGTTTTCGCGGCAGCGGGGTCTGGCCGCG 660
              ||||| |||||| | ||     | ||||||||||    |||||||| || || || | |
Db      20905 ACGGTCATGTCGACGCCCACGACGTTCGTGGAGTTCAGCCGGCAGCGCGGGCTCGCGGAG 20964

Qy        661 GACGGGCATTGCAAGGCGTTCTCGGCGGCGGCGGACGGGACCGGCTGGGGTGAGGGTGTG 720
              ||||| |  | ||||||||| | ||| ||||||||||   |||| ||  ||||| ||
Db      20965 GACGGCCGGTCGAAGGCGTTCGCCGCGTCGGCGGACGGCTTCGGCCCGGCGGAGGGCGTC 21024

Qy        721 GGGATGCTGCTGGTGGAGCGGCTCTCCGACGCCCATCGCAACGGTCACCGTGTCCTGGCC 780
              || ||||| || || ||||| || || |||||||  |||||||| ||||||||  |||||
Db      21025 GGCATGCTCCTCGTCGAGCGCCTGTCGGACGCCCGCCGCAACGGACACCGTGTGCTGGCG 21084

Qy        781 GTGGTGCGTGGCAGTGCGGTCAACCAGGACGGTGCGAGCAACGGTCTGACCGCGCCCAAC 840
              || ||||| ||||| ||||||||||||||||| |||||||||| |||||||| || |||
Db      21085 GTCGTGCGCGGCAGCGCGGTCAACCAGGACGGCGCGAGCAACGGCCTGACCGCCCCGAAC 21144

Qy        841 GGGCCGTCCCAGCAGCGTGTCATCCGCCAGGCCCTCGCCAACGCCGGCTTGTCGGCCGGT 900
              ||||| || |||||||| |||||||| |  || |||||  |||| |  || || |||
Db      21145 GGGCCCTCGCAGCAGCGCGTCATCCGGCGCGCGCTCGCGGACGCCCGACTGACGACCGCC 21204
```

```
Qy        901 GATGTCGACGCGGTGGAGGCCCACGGCACCGGCACCACTTTGGGCGACCCGATCGAGGCC 960
Db      21205 GACGTGGACGTCGTCGAGGCCCACGGCACGGGCACGCGACTCGGCGACCCGATCGAGGCA 21264

Qy        961 CAGGCCCTCCTCGCGACCTACGGACAGGACCGTGCCGGCGAGGGGCCGCTGTGGCTGGGC 1020
Db      21265 CAGGCCCTCATCGCCACCTACGGCCAGGGGCGCGACACCGAACAGCCGCTGCGCCTGGGG 21324

Qy       1021 TCGGTCAAGTCCAATGTCGGTCACACACAGGCTGCCGCGGGCGTCGCCGGGGTGATCAAG 1080
Db      21325 TCGTTGAAGTCCAACATCGGACACACCCAGGCCGCCGCCGGTGTCTCCGGCATCATCAAG 21384

Qy       1081 ATGGTGATGGCGCTGCGGCATGGTCTGCTGCCGCGGACGTTGCATGTGGATGAGCCGTCG 1140
Db      21385 ATGGTCCAGGCGATGCGCCACGGCGTCCTGCCGAAGACGCTCCACGTGGACCGGCCGTCG 21444

Qy       1141 CCGCATGTGGACTGGTCCGCGGGTGCGGTGCAGCTGCTGACGGAGACGGTGCCCTGGCCC 1200
Db      21445 GACCAGATCGACTGGTCGGCGGGCACGGTCGAGCTGCTCACCGAGGCCATGGACTGGCCG 21504

Qy       1201 GG---CGGGGAGGGGCGGCTACGGCGGGCAGGAGTGTCATCATTCGGCGTCAGCGGCACC 1257
Db      21505 AGGAAGCAGGAGGGCGGGCTGCGCCGCGCGGCCGTCTCCTCCTTCGGCATCAGCGGCACG 21564

Qy       1258 AACGCCCACGTCATCCTCGAAGAAGCACCCGCC 1290
Db      21565 AACGCGCACATCGTGCTCGAAGAAGCCCCGGTC 21597
```

RESULT 14
AAZ87285
ID AAZ87285 standard; DNA; 37948 BP.
XX
AC AAZ87285;
XX
DT 05-JUN-2000 (first entry)
XX
DE S. venezuelae pik (macrolide biosynthesis) gene cluster, SEQ ID NO:5.
XX
KW Desosamine biosynthesis; macrolide; polyketide; methymycin; pikromycin;
KW neomethymycin; narbomycin; polyhydroxyalkanoate monomer synthase;
KW biopolymer; antibiotic; chemotherapeutic; immunosuppressant; asthma;
KW chronic obstructive pulmonary disease; respiratory inflammation;
KW hypercholesterolaemia; crop protection agent; ds.
XX
OS Streptomyces venezuelae ATCC15439.
XX
PN WO200000620-A2.
XX
PD 06-JAN-2000.
XX
PF 25-JUN-1999; 99WO-US14398.
XX
PR 26-JUN-1998; 98US-0105537.
XX
PA (MINU) UNIV MINNESOTA.
XX
PI Sherman DH, Liu H, Xue Y, Zhao L;
XX
DR WPI; 2000-160679/14.
DR P-PSDB; AAY77180.
XX
PT Desosamine and macrolide biosynthetic gene clusters, useful for, e .g.
PT synthesis of methymycin and pikromycin -
XX
PS Claim 13; Page 299-315; 438pp; English.
XX
CC The invention relates to an isolated and purified nucleic acid segment
CC comprising a desosamine biosynthetic gene cluster, a fragment or its
CC biologically active variant, where the nucleic acid sequence is not
CC derived from the eryC gene cluster of Saccharopolyspora erythraea or
CC Streptomyces antibioticus. The invention also relates to a macrolide
CC biosynthetic gene cluster, or fragments thereof. The macrolide
CC biosynthetic gene cluster encodes proteins which synthesise methymycin,
CC pikromycin, neomethymycin, narbomycin or a combination of these
CC compounds. Recombinant or augmented cells comprising the desosamine
CC and/or macrolide biosynthetic gene clusters are useful for the production
CC of biologically active macrolides. The macrolide biosynthetic proteins
CC are useful for synthesis of methymycin, pikromycin, neomethymycin and
CC narbomycin. The alternative termination of polyketide synthesis may be
CC useful to prepare novel antibiotics and polyhydroxyalkanoate (PHA)
CC monomers. The compounds produced by the recombinant host cells are useful
CC as biopolymers, e.g., in packaging or biomedical applications, to
CC engineer PHA monomer synthases or to prepare biologically active agents,
CC such as chemotherapeutics, immunosuppressants, agents to treat asthma,
CC chronic obstructive pulmonary disease as well as other diseases involving
CC respiratory inflammation, cholesterol-lowering agents or macrolide-based
CC antibiotics which are active against a variety of organisms, e.g.,
CC bacteria, including multi-drug resistant pneumococci and other
CC respiratory pathogens, as well as viral parasitic pathogens, or as crop
CC protection agents (e.g., fungicides or insecticides) via expression of
CC polyketides in plants. The present sequence represents the macrolide
CC biosynthetic gene cluster (pik) from Streptomyces venezuelae ATCC 15439.
XX
SQ Sequence 37948 BP; 4901 A; 14718 C; 13311 G; 5018 T; 0 other;

Query Match 55.5%; Score 715.4; DB 21; Length 37948;
Best Local Similarity 72.6%; Pred. No. 6.5e-111;
Matches 939; Conservative 0; Mismatches 351; Indels 3; Gaps 1;

```
Qy          1 GAGCCGATCGCCATCGTTGGCATGGCCTGTCGTTTCCCCGGCGGAGTGACCTCGGCGGAC 60
Db      21475 GAGCCCGTCGCGATCGTCGGCATGGCGTGCCGCCTGCCGGGCGGTGTCGCCTCGCCGGAG 21534

Qy         61 GACTTCTGGGATCTGATCTCCTCCGAGCAGGACGCGATCGGCGGATTCCCCACCGACCGC 120
Db      21535 GACCTGTGGCGGCTGGTGGCCGGCGGCGAGGACGCGATCTCCGGCTTCCCGCAGGACCGC 21594

Qy        121 GGCTGGGACCTGGACACGCTCTACGACCCCGACCCCGACCACCCCGGCACCTGCTACACC 180
Db      21595 GGCTGGGACGTGGAGGGGCTGTACGACCCGGACCCGGACGCGTCCGGGCGGACGTACTGC 21654

Qy        181 CGAAACGGCGGATTCCTCTACGACGCAGGCCACTTCGACGCCGAATTCTTCGGCATCAGC 240
Db      21655 CGTGCCGGTGGCTTCCTCGACGAGGCGGGCGAGTTCGACGCCGACTTCTTCGGGATCTCG 21714

Qy        241 CCCCGCGAAGCCCTCGCCATGGACCCCCAGCAACGACTCCTCCTCGAAACCGCCTGGGAA 300
Db      21715 CCGCGCGAGGCCCTCGCCATGGACCCGCAGCAGCGGCTCCTCCTGGAGACCTCCTGGGAG 21774

Qy        301 ACCATCGAACACGCCGGCATCAACCCCCACACCCTCCACGGCACCCCCACCGGAGTCTTC 360
Db      21775 GCCGTCGAGGACGCCGGGATCGACCCGACCTCCCTTCAGGGGCAGCAGGTCGGCGTGTTC 21834

Qy        361 ACCGGCACCAACGGACAGGACTACGCACTTCGCGTGCACAACGCGGGCCAGTCAACCGAT 420
Db      21835 GCGGGCACCAACGGCCCCCACTACGAGCCGCTGCTCCGCAACACCGCCGAGGATCTTGAG 21894

Qy        421 GGTTTCGCACTGACCGGAACCGCCGGCAGCGTCATCTCCGGTCGTATCTCGTACACGTTT 480
Db      21895 GGTTACGTCGGGACGGGCAACGCCGCCAGCATCATGTCGGGCCGTGTCTCGTACACCCTC 21954

Qy        481 GGTTTTGAGGGTCCTGCGGTGTCGGTGGACACGGCTTGTTCCTCGTCGTTGGTGGCTTTG 540
Db      21955 GGCCTGGAGGGCCCGGCCGTCACGGTCGACACCGCCTGCTCCTCCTCGCTGGTCGCCCTG 22014

Qy        541 CATCTGGCCTGTCAGGCGTTGCGTGCGGGTGAGTGCTCGATGGCGCTTGCCGGGGGTGTG 600
Db      22015 CACCTCGCCGTGCAGGCCCTGCGCAAGGGCGAATGCGGACTGGCGCTCGCGGGCGGTGTG 22074

Qy        601 ACGGTGATGTCGTCTCCGGGTGCCTTCGTGGAGTTTTCGCGGCAGCGGGGTCTGGCCGCG 660
Db      22075 ACGGTCATGTCGACGCCCACGACGTTCGTGGAGTTCAGCCGGCAGCGCGGGCTCGCGGAG 22134

Qy        661 GACGGGCATTGCAAGGCGTTCTCGGCGGCGGCGGACGGGACCGGCTGGGGTGAGGGTGTG 720
```

```
Db      22135 GACGGCCGGTCGAAGGCGTTCGCCGCGTCGGCGGACGGCTTCGGCCCGGCGGAGGGCGTC 22194

Qy        721 GGGATGCTGCTGGTGGAGCGGCTCTCCGACGCCCATCGCAACGGTCACCGTGTCCTGGCC 780

Db      22195 GGCATGCTCCTCGTCGAGCGCCTGTCGGACGCCCGCCGCAACGGACACCGTGTGCTGGCG 22254

Qy        781 GTGGTGCGTGGCAGTGCGGTCAACCAGGACGGTGCGAGCAACGGTCTGACCGCGCCCAAC 840

Db      22255 GTCGTGCGCGGCAGCGCGGTCAACCAGGACGGCGCGAGCAACGGCCTGACCGCCCCGAAC 22314

Qy        841 GGGCCGTCCCAGCAGCGTGTCATCCGCCAGGCCCTCGCCAACGCCGGCTTGTCGGCCGGT 900

Db      22315 GGGCCCTCGCAGCAGCGCGTCATCCGGCGCGCGCTCGCGGACGCCCGACTGACGACCGCC 22374

Qy        901 GATGTCGACGCGGTGGAGGCCCACGGCACCGGCACCACTTTGGGCGACCCGATCGAGGCC 960

Db      22375 GACGTGGACGTCGTCGAGGCCCACGGCACGGGCACGCGACTCGGCGACCCGATCGAGGCA 22434

Qy        961 CAGGCCCTCCTCGCGACCTACGGACAGGACCGTGCCGGCGAGGGGCCGCTGTGGCTGGGC 1020

Db      22435 CAGGCCCTCATCGCCACCTACGGCCAGGGGCGCGACACCGAACAGCCGCTGCGCCTGGGG 22494

Qy       1021 TCGGTCAAGTCCAATGTCGGTCACACACAGGCTGCCGCGGGCGTCGCCGGGGTGATCAAG 1080

Db      22495 TCGTTGAAGTCCAACATCGGACACACCCAGGCCGCCGCCGGTGTCTCCGGCATCATCAAG 22554

Qy       1081 ATGGTGATGGCGCTGCGGCATGGTCTGCTGCCGCGGACGTTGCATGTGGATGAGCCGTCG 1140

Db      22555 ATGGTCCAGGCGATGCGCCACGGCGTCCTGCCGAAGACGCTCCACGTGGACCGGCCGTCG 22614

Qy       1141 CCGCATGTGGACTGGTCCGCGGGTGCGGTGCAGCTGCTGACGGAGACGGTGCCCTGGCCC 1200

Db      22615 GACCAGATCGACTGGTCGGCGGGCACGGTCGAGCTGCTCACCGAGGCCATGGACTGGCCG 22674

Qy       1201 GG---CGGGGAGGGGCGGCTACGGCGGGCAGGAGTGTCATCATTCGGCGTCAGCGGCACC 1257

Db      22675 AGGAAGCAGGAGGGCGGGCTGCGCCGCGCGGCCGTCTCCTCCTTCGGCATCAGCGGCACG 22734

Qy       1258 AACGCCCACGTCATCCTCGAAGAAGCACCCGCC 1290

Db      22735 AACGCGCACATCGTGCTCGAAGAAGCCCCGGTC 22767


RESULT 15
AAA75633
ID   AAA75633 standard; DNA; 38506 BP.
XX
AC   AAA75633;
XX
DT   22-JAN-2001  (first entry)
XX
DE   Nucleotide sequence of the insert DNA in cosmid pKOS023-27.
XX
KW   Narbonolide synthase; polyketide synthase gene; narbonolide polyketide;
KW   antibiotic; C12-hydroxyalse; picK; desosamine biosynthesis;
KW   desosaminyl transferase enzyme; ketolide; beta-glucosidase enzyme;
KW   picromycin biosynthesis; ss.
XX
OS   Streptomyces venezuelae.
XX
PN   US6117659-A.
XX
PD   12-SEP-2000.
XX
PF   27-MAY-1999;   99US-0320878.
XX
PR   28-MAY-1998;   98US-0087080.
PR   22-SEP-1998;   98US-0100880.
PR   08-FEB-1999;   99US-0119139.
PR   20-MAY-1999;   99US-0134990.
PR   30-APR-1997;   97US-0846247.
PR   06-MAY-1998;   98US-0073538.
PR   28-AUG-1998;   98US-0141908.
XX
PA   (KOSA-) KOSAN BIOSCIENCES INC.
XX
PI   Ashley G,  Betlach MC,  Betlach M,  Tang L,  McDaniel R;
XX
DR   WPI; 2000-610844/58.
XX
PT   New recombinant picK hydroxylase gene of Streptomyces venezuelae useful
PT   for converting ketolides to antibiotics and as antibiotics and
PT   intermediates in the synthesis of compounds with pharmaceutical value
PT   -
XX
PS   Disclosure; Columns 15-32; 117pp; English.
XX
CC   The present sequence is used to produce the recombinant DNA compounds
CC   of the invention. The specification describes a recombinant DNA compound
CC   expressing recombinant polyketide synthase genes in host cells for the
CC   production of narbonolide, narbonolide derivatives and polyketides that
CC   are useful as antibiotics and as intermediates in the synthesis of
CC   compounds with pharmaceutical value. The DNA compounds may also encode
CC   a C12-hydroxyalse (picK), desosamine biosynthesis and desosaminyl
CC   transferase enzymes (useful for conversion of ketolides to antibiotics),
CC   and the beta-glucosidase enzyme (involved in picromycin biosynthesis).
CC   These compounds are also useful for increasing the antibiotic activity
CC   of a compound relative to the unhydroxylated compound. The recombinant
CC   host cells are useful as genetic systems that allow rapid engineering
CC   of the narbonolide polyketide synthase. These would be valuable for
CC   creating novel ketolide analogs for pharmaceutical applications.
XX
SQ   Sequence 38506 BP; 4914 A; 15118 C; 13444 G; 5030 T; 0 other;

  Query Match             55.5%;  Score 715.4;  DB 21;  Length 38506;
  Best Local Similarity   72.6%;  Pred. No. 6.5e-111;
  Matches  939;  Conservative    0;  Mismatches  351;  Indels    3;  Gaps    1;

Qy          1 GAGCCGATCGCCATCGTTGGCATGGCCTGTCGTTTCCCCGGCGGAGTGACCTCGGCGGAC 60

Db      18447 GAGCCCGTCGCGATCGTCGGCATGGCGTGCCGCCTGCCGGGCGGTGTCGCCTCGCCGGAG 18506

Qy         61 GACTTCTGGGATCTGATCTCCTCCGAGCAGGACGCGATCGGCGGATTCCCCACCGACCGC 120

Db      18507 GACCTGTGGCGGCTGGTGGCCGGCGGCGAGGACGCGATCTCCGGCTTCCCGCAGGACCGC 18566

Qy        121 GGCTGGGACCTGGACACGCTCTACGACCCCGACCCCGACCACCCCGGCACCTGCTACACC 180

Db      18567 GGCTGGGACGTGGAGGGGCTGTACGACCCGGACCCGGACGCGTCCGGGCGGACGTACTGC 18626

Qy        181 CGAAACGGCGGATTCCTCTACGACGCAGGCCACTTCGACGCCGAATTCTTCGGCATCAGC 240

Db      18627 CGTGCCGGTGGCTTCCTCGACGAGGCGGGCGAGTTCGACGCCGACTTCTTCGGGATCTCG 18686

Qy        241 CCCCGCGAAGCCCTCGCCATGGACCCCCAGCAACGACTCCTCCTCGAAACCGCCTGGGAA 300

Db      18687 CCGCGCGAGGCCCTCGCCATGGACCCGCAGCAGCGGCTCCTCCTGGAGACCTCCTGGGAG 18746

Qy        301 ACCATCGAACACGCCGGCATCAACCCCCACACCCTCCACGGCACCCCCACCGGAGTCTTC 360

Db      18747 GCCGTCGAGGACGCCGGGATCGACCCGACCTCCCTTCAGGGGCAGCAGGTCGGCGTGTTC 18806

Qy        361 ACCGGCACCAACGGACAGGACTACGCACTTCGCGTGCACAACGCGGGCCAGTCAACCGAT 420

Db      18807 GCGGGCACCAACGGCCCCCACTACGAGCCGCTGCTCCGCAACACCGCCGAGGATCTTGAG 18866

Qy        421 GGTTTCGCACTGACCGGAACCGCCGGCAGCGTCATCTCCGGTCGTATCTCGTACACGTTT 480

Db      18867 GGTTACGTCGGGACGGGCAACGCCGCCAGCATCATGTCGGGCCGTGTCTCGTACACCCTC 18926

Qy        481 GGTTTTGAGGGTCCTGCGGTGTCGGTGGACACGGCTTGTTCCTCGTCGTTGGTGGCTTTG 540

Db      18927 GGCCTGGAGGGCCCGGCCGTCACGGTCGACACCGCCTGCTCCTCCTCGCTGGTCGCCCTG 18986

Qy        541 CATCTGGCCTGTCAGGCGTTGCGTGCGGGTGAGTGCTCGATGGCGCTTGCCGGGGGTGTG 600
```

```
Db      18987 CACCTCGCCGTGCAGGCCCTGCGCAAGGGCGAATGCGGACTGGCGCTCGCGGGCGGTGTG 19046

Qy        601 ACGGTGATGTCGTCTCCGGGTGCCTTCGTGGAGTTTTCGCGGCAGCGGGGTCTGGCCGCG 660
              ||||| |||||| | ||     | ||||||||||    |||||||| || || || | |
Db      19047 ACGGTCATGTCGACGCCCACGACGTTCGTGGAGTTCAGCCGGCAGCGCGGGCTCGCGGAG 19106

Qy        661 GACGGGCATTGCAAGGCGTTCTCGGCGGCGGCGGACGGGACCGGCTGGGGTGAGGGTGTG 720
              ||||| |   |  |||||||||| | ||| ||||||||||    ||||  ||  ||||| ||
Db      19107 GACGGCCGGTCGAAGGCGTTCGCCGCGTCGGCGGACGGCTTCGGCCCGGCGGAGGGCGTC 19166

Qy        721 GGGATGCTGCTGGTGGAGCGGCTCTCCGACGCCCATCGCAACGGTCACCGTGTCCTGGCC 780
              || ||||| || || ||||| || || |||||||  |||||||| |||||||| |||||
Db      19167 GGCATGCTCCTCGTCGAGCGCCTGTCGGACGCCCGCCGCAACGGACACCGTGTGCTGGCG 19226

Qy        781 GTGGTGCGTGGCAGTGCGGTCAACCAGGACGGTGCGAGCAACGGTCTGACCGCGCCCAAC 840
              || ||||| ||||| |||||||||||||||||| |||||||||| |||||||| || |||
Db      19227 GTCGTGCGCGGCAGCGCGGTCAACCAGGACGGCGCGAGCAACGGCCTGACCGCCCCGAAC 19286

Qy        841 GGGCCGTCCCAGCAGCGTGTCATCCGCCAGGCCCTCGCCAACGCCGGCTTGTCGGCCGGT 900
              ||||| || |||||||| |||||||| |  || |||||  |||||  |  || || |||
Db      19287 GGGCCCTCGCAGCAGCGCGTCATCCGGCGCGCGCTCGCGGACGCCCGACTGACGACCGCC 19346

Qy        901 GATGTCGACGCGGTGGAGGCCCACGGCACCGGCACCACTTTGGGCGACCCGATCGAGGCC 960
              || || ||||  || |||||||||||||| |||||     | ||||||||||||||||||
Db      19347 GACGTGGACGTCGTCGAGGCCCACGGCACGGGCACGCGACTCGGCGACCCGATCGAGGCA 19406

Qy        961 CAGGCCCTCCTCGCGACCTACGGACAGGACCGTGCCGGCGAGGGGCCGCTGTGGCTGGGC 1020
              ||||||||| |||| |||||||| ||||  || | |  |||   ||||||| | |||||
Db      19407 CAGGCCCTCATCGCCACCTACGGCCAGGGGCGCGACACCGAACAGCCGCTGCGCCTGGGG 19466

Qy       1021 TCGGTCAAGTCCAATGTCGGTCACACACAGGCTGCCGCGGGCGTCGCCGGGGTGATCAAG 1080
              ||| | ||||||||  |||| ||||| ||||| ||||| || ||| ||||  | ||||||
Db      19467 TCGTTGAAGTCCAACATCGGACACACCCAGGCCGCCGCCGGTGTCTCCGGCATCATCAAG 19526

Qy       1081 ATGGTGATGGCGCTGCGGCATGGTCTGCTGCCGCGGACGTTGCATGTGGATGAGCCGTCG 1140
              |||||   |||| |||| || ||  | ||||||  |||| | || |||||   |||||||
Db      19527 ATGGTCCAGGCGATGCGCCACGGCGTCCTGCCGAAGACGCTCCACGTGGACCGGCCGTCG 19586

Qy       1141 CCGCATGTGGACTGGTCCGCGGGTGCGGTGCAGCTGCTGACGGAGACGGTGCCCTGGCCC 1200
                ||  | |||||||| |||||  ||||  ||||||| || ||| |  ||  ||||||
Db      19587 GACCAGATCGACTGGTCGGCGGGCACGGTCGAGCTGCTCACCGAGGCCATGGACTGGCCG 19646

Qy       1201 GG---CGGGGAGGGGCGGCTACGGCGGGCAGGAGTGTCATCATTCGGCGTCAGCGGCACC 1257
               |       ||||||  |||| || || || |  || || || |||||| ||||||||||
Db      19647 AGGAAGCAGGAGGGCGGGCTGCGCCGCGCGGCCGTCTCCTCCTTCGGCATCAGCGGCACG 19706

Qy       1258 AACGCCCACGTCATCCTCGAAGAAGCACCCGCC 1290
              ||||| ||| || | |||||||||||| || | |
Db      19707 AACGCGCACATCGTGCTCGAAGAAGCCCCGGTC 19739

Search completed: June 17, 2003, 13:48:35
Job time : 261.707 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 17, 2003, 11:45:13 ; Search time 55.8814 Seconds
(without alignments)
7079.508 Million cell updates/sec

Title: US-09-914-286-1_COPY_6256_7545
Perfect score: 1290
Sequence: 1 gagccgatcgccatcgttgg.....tcctcgaagaagcacccgcc 1290

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	732.2	56.8	44377	2	US-08-804-227C-7	Sequence 7, Appli
	2	732.2	56.8	44377	2	US-08-804-198-1	Sequence 1, Appli
	3	722.6	56.0	13987	2	US-08-804-227C-13	Sequence 13, Appl
	4	715.4	55.5	11220	4	US-09-105-537-32	Sequence 32, Appl
	5	715.4	55.5	36778	4	US-09-105-537-5	Sequence 5, Appli
	6	715.4	55.5	38506	3	US-09-320-878-19	Sequence 19, Appl
	7	715.2	55.4	80161	3	US-09-036-987A-1	Sequence 1, Appli
	8	715.2	55.4	80161	4	US-09-370-700-1	Sequence 1, Appli
	9	689.4	53.4	13842	4	US-09-105-537-30	Sequence 30, Appl
	10	688.4	53.4	43280	2	US-08-804-227C-1	Sequence 1, Appli
	11	684.6	53.1	20235	1	US-07-642-734C-3	Sequence 3, Appli
	12	684.6	53.1	20235	3	US-08-439-009A-3	Sequence 3, Appli
	13	680.4	52.7	15872	4	US-09-105-537-1	Sequence 1, Appli
	14	673	52.2	4041	4	US-09-105-537-36	Sequence 36, Appl
	15	673	52.2	50937	4	US-09-428-517-1	Sequence 1, Appli
	16	655.2	50.8	4689	4	US-09-105-537-34	Sequence 34, Appl
	17	624	48.4	11219	1	US-07-642-734C-1	Sequence 1, Appli
	18	624	48.4	11219	3	US-08-439-009A-1	Sequence 1, Appli
c	19	608.8	47.2	4411529	4	US-09-103-840A-1	Sequence 1, Appli
c	20	596.8	46.3	4403765	4	US-09-103-840A-2	Sequence 2, Appli
	21	571.6	44.3	28958	1	US-08-258-261B-6	Sequence 6, Appli
	22	571.6	44.3	28958	1	US-08-456-837-6	Sequence 6, Appli
	23	571.6	44.3	28958	1	US-08-457-342-6	Sequence 6, Appli
	24	571.6	44.3	28958	1	US-08-457-646A-6	Sequence 6, Appli
	25	571.6	44.3	28958	1	US-08-458-076A-6	Sequence 6, Appli
	26	571.6	44.3	28958	1	US-08-764-233A-4	Sequence 4, Appli
	27	571.6	44.3	28958	1	US-08-457-335A-6	Sequence 6, Appli
	28	571.6	44.3	28958	1	US-08-729-214-6	Sequence 6, Appli
	29	571.6	44.3	28958	3	US-09-028-934-6	Sequence 6, Appli
	30	571.6	44.3	49377	1	US-08-764-233A-1	Sequence 1, Appli
	31	565.6	43.8	33529	4	US-09-144-085-3	Sequence 3, Appli
	32	553	42.9	4403765	4	US-09-103-840A-2	Sequence 2, Appli
	33	553	42.9	4411529	4	US-09-103-840A-1	Sequence 1, Appli
	34	522.6	40.5	68750	3	US-09-335-409-1	Sequence 1, Appli
	35	522.6	40.5	68750	4	US-09-568-102-1	Sequence 1, Appli
	36	522.6	40.5	68750	4	US-09-567-969-1	Sequence 1, Appli
	37	522.6	40.5	68750	4	US-09-568-480-1	Sequence 1, Appli
	38	522.6	40.5	68750	4	US-09-568-486-1	Sequence 1, Appli
	39	522.6	40.5	68750	4	US-09-568-472-1	Sequence 1, Appli
	40	522.6	40.5	68750	4	US-09-567-899-1	Sequence 1, Appli
	41	519.4	40.3	71989	4	US-09-443-501A-2	Sequence 2, Appli
c	42	504	39.1	1419	4	US-09-434-288-2	Sequence 2, Appli
	43	454	35.2	1462	4	US-09-434-288-4	Sequence 4, Appli
	44	340.6	26.4	751	3	US-09-010-809-4	Sequence 4, Appli
c	45	337.4	26.2	1434	4	US-09-434-288-3	Sequence 3, Appli

ALIGNMENTS

```
RESULT 1
US-08-804-227C-7
; Sequence 7, Application US/08804227C
; Patent No. 5876991
;  GENERAL INFORMATION:
;    APPLICANT:  DeHoff, Bradley S.
;    APPLICANT:  Kuhstoss, Stuart A.
;    APPLICANT:  Rosteck, Paul R., Jr.
;    APPLICANT:  Sutton, Kimberly L.
;    TITLE OF INVENTION:  POLYKETIDE SYNTHASE GENES
;    NUMBER OF SEQUENCES:  15
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  THOMAS G. PLANT  1501
;      STREET:  LILLY CORPORATE CENTER
;      CITY:  INDIANAPOLIS
;      STATE:  IN
;      COUNTRY:  USA
;      ZIP:  46285
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  MS-DOS
;      SOFTWARE:  ASCI(DOS) Text only
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/804,227C
;      FILING DATE:  February 21, 1997
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Plant, Thomas, G.
;      REGISTRATION NUMBER:  35,784
;      REFERENCE/DOCKET NUMBER:  X-8231
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  317-276-2459
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  44377 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  350..14002
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  14046..20036
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  20110..31284
;    FEATURE:
```

```
;      NAME/KEY:  CDS
;      LOCATION:  31329..36071
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  36155..41830
US-08-804-227C-7

  Query Match             56.8%;  Score 732.2;  DB 2;  Length 44377;
  Best Local Similarity   73.0%;  Pred. No. 4.7e-144;
  Matches  941;  Conservative    0;  Mismatches  348;  Indels    0;  Gaps    0;

Qy          1 GAGCCGATCGCCATCGTTGGCATGGCCTGTCGTTTCCCCGGCGGAGTGACCTCGGCGGAC 60
              |||||||||||| ||||| |||||||| || |||| ||| ||||| ||| | ||| |||||
Db      14148 GAGCCGATCGCGATCGTGGGCATGGCGTGCCGTTACCCGGGCGGTGTGGCGTCGCCGGAC 14207

Qy         61 GACTTCTGGGATCTGATCTCCTCCGAGCAGGACGCGATCGGCGGATTCCCCACCGACCGC 120
              ||| | ||||| ||| |   |  |||    ||||||| ||  ||  ||||||  ||||||
Db      14208 GACCTGTGGGACCTGGTGGCAGCCGGTACGGACGCGGTCTCCGCGTTCCCCGTCGACCGT 14267

Qy        121 GGCTGGGACCTGGACACGCTCTACGACCCCGACCCCGACCACCCCGGCACCTGCTACACC 180
              ||||||||| | ||   ||| |||||||||||| || ||         ||      | |||
Db      14268 GGCTGGGACGTCGAGGGGCTGTACGACCCCGATCCGGAGGCGGTGGGGCGTAGTTACGTG 14327

Qy        181 CGAAACGGCGGATTCCTCTACGACGCAGGCCACTTCGACGCCGAATTCTTCGGCATCAGC 240
              ||  | ||||| |||||  ||   || | | | |||||||| || |||||||| |||
Db      14328 CGGGAGGGCGGGTTCCTGCACTCGGCGGCCGAGTTCGACGCGGAGTTCTTCGGGATCTCG 14387

Qy        241 CCCCGCGAAGCCCTCGCCATGGACCCCCAGCAACGACTCCTCCTCGAAACCGCCTGGGAA 300
              |||||  || ||    || ||||| || ||||| ||  | || || || ||  | |||||
Db      14388 CCCCGTGAGGCGGCGGCGATGGATCCGCAGCAGCGGTTGCTGCTGGAGACGTCGTGGGAG 14447

Qy        301 ACCATCGAACACGCCGGCATCAACCCCCACACCCTCCACGGCACCCCCACCGGAGTCTTC 360
              |  | || |  || || |||  ||||    | || | ||||||||  ||||| ||||||
Db      14448 GCGCTGGAGCGGGCGGGGATCGTCCCCGCGTCGCTGCGCGGCACCCGTACCGGCGTCTTC 14507

Qy        361 ACCGGCACCAACGGACAGGACTACGCACTTCGCGTGCACAACGCGGGCCAGTCAACCGAT 420
              ||||||  ||      | |||||||     ||  |  ||   ||    | |     |||
Db      14508 ACCGGCGTCATGTACGACGACTACGGGTCGCGGTTCGACTCGGCTCCGCCGGAGTACGAG 14567

Qy        421 GGTTTCGCACTGACCGGAACCGCCGGCAGCGTCATCTCCGGTCGTATCTCGTACACGTTT 480
              || | |   ||| ||| | |||||||||| ||   |||||||  |   | ||  ||||
Db      14568 GGCTACCTCGTGAACGGCAGCGCCGGCAGCATCGCGTCCGGTCGGGTTGCCTATGCGTTG 14627

Qy        481 GGTTTTGAGGGTCCTGCGGTGTCGGTGGACACGGCTTGTTCCTCGTCGTTGGTGGCTTTG 540
              || || ||||| || ||| || |||||||||||||| ||||| ||||||||||||||| |||
Db      14628 GGGTTGGAGGGGCCGGCGCTGACGGTGGACACGGCGTGTTCGTCGTCGTTGGTGGCGTTG 14687

Qy        541 CATCTGGCCTGTCAGGCGTTGCGTGCGGGTGAGTGCTCGATGGCGCTTGCCGGGGGTGTG 600
              |||||||    ||| |||||||   |||||||||     ||||| | ||||| || |||
Db      14688 CATCTGGCGGTGCAGTCGTTGCGGCGGGGTGAGTGTGATCTGGCGTTGGCCGGTGGGGTG 14747

Qy        601 ACGGTGATGTCGTCTCCGGGTGCCTTCGTGGAGTTTTCGCGGCAGCGGGGTCTGGCCGCG 660
              |||||||||| || | |||    |   |||||||||| |||||||||||||| ||||| |||
Db      14748 ACGGTGATGGCGACGCCGACGGTGCTCGTGGAGTTCTCGCGGCAGCGGGGGCTGGCGGCG 14807

Qy        661 GACGGGCATTGCAAGGCGTTCTCGGCGGCGGCGGACGGGACCGGCTGGGGTGAGGGTGTG 720
              |||||||  |||||||||||| ||| ||  ||||||||||| |   ||||  |||||||||
Db      14808 GACGGGCGGTGCAAGGCGTTCGCGGAGGGTGCGGACGGGACGGCGTGGGCCGAGGGTGTG 14867

Qy        721 GGGATGCTGCTGGTGGAGCGGCTCTCCGACGCCCATCGCAACGGTCACCGTGTCCTGGCC 780
              ||  |||||||||||||||||||||||||||||  ||||| || || || || |||||
Db      14868 GGCGTGCTGCTGGTGGAGCGGCTCTCCGACGCCCGCCGCGCAATGGCCATCGGGTGCTGGCG 14927

Qy        781 GTGGTGCGTGGCAGTGCGGTCAACCAGGACGGTGCGAGCAACGGTCTGACCGCGCCCAAC 840
              |||||||| |||||||||||||| |||||||||||||||||||| ||||| ||||| |
Db      14928 GTGGTGCGGGGCAGTGCGGTCAATCAGGACGGTGCGAGCAACGGGCTGACGGCGCCGAGT 14987

Qy        841 GGGCCGTCCCAGCAGCGTGTCATCCGCCAGGCCCTCGCCAACGCCGGCTTGTCGGCCGGT 900
              ||  ||  | |||||||| || |||||  |||| || ||| |||| ||  || || |||
Db      14988 GGTCCTGCGCAGCAGCGGGTGATCCGTGAGGCGCTGGCCGACGCGGGGCTGACGCCCGCC 15047

Qy        901 GATGTCGACGCGGTGGAGGCCCACGGCACCGGCACCACTTTGGGCGACCCGATCGAGGCC 960
              ||  |||||||||||| ||||| |||||||||||||| |  |||||||||| ||||||||

Db      15048 GACGTCGACGCGGTCGAGGCGCACGGCACCGGCACACCCCTGGGCGACCCCATCGAGGCG 15107

Qy        961 CAGGCCCTCCTCGCGACCTACGGACAGGACCGTGCCGGCGAGGGGCCGCTGTGGCTGGGC 1020
              ||  | || || ||||| ||    || ||    ||| | || ||| ||||| ||||
Db      15108 GGTGCGTTGCTGGCCACCTATGGCAGTGAGCGCCAGGGCCAAGGTCCGTTGTGGTTGGGG 15167

Qy       1021 TCGGTCAAGTCCAATGTCGGTCACACACAGGCTGCCGCGGGCGTCGCCGGGGTGATCAAG 1080
              ||| | ||||| ||  |||| ||  | ||||| || ||||| || |  || |||||||||
Db      15168 TCGTTGAAGTCGAACATCGGGCATGCGCAGGCGGCTGCGGGTGTGGGTGGCGTGATCAAG 15227

Qy       1081 ATGGTGATGGCGCTGCGGCATGGTCTGCTGCCGCGGACGTTGCATGTGGATGAGCCGTCG 1140
              |||||  |||| |||||||||    | |||||||||| |||||||||| |||||||
Db      15228 GTGGTGCAGGCGATGCGGCATGGGTCGTTGCCGCGGACGCTGCATGTGGATGCGCCGTCG 15287

Qy       1141 CCGCATGTGGACTGGTCCGCGGGTGCGGTGCAGCTGCTGACGGAGACGGTGCCCTGGCCC 1200
              || | ||||| ||| |  |||||||||| |||||||||| |||||   | | |||||
Db      15288 TCGAAGGTGGAGTGGGCTTCGGGTGCGGTGGAGCTGCTGACCGAGACCCGGTCGTGGCCG 15347

Qy       1201 GGCGGGGAGGGGCGGCTACGGCGGGCAGGAGTGTCATCATTCGGCGTCAGCGGCACCAAC 1260
              |  ||| || |||| | |||||||| |  |||||  | ||||| || ||||| ||||||
Db      15348 CGGCGGGTGGAGCGGGTGCGGCGGGCCGCGGTGTCGGCGTTCGGGGTGAGCGGGACCAAC 15407

Qy       1261 GCCCACGTCATCCTCGAAGAAGCACCCGC 1289
              ||||| ||  |||| || ||||| || ||
Db      15408 GCCCATGTGGTCCTGGAGGAAGCGCCGGC 15436


RESULT 2
US-08-804-198-1
; Sequence 1, Application US/08804198
; Patent No. 5945320
;  GENERAL INFORMATION:
;    APPLICANT:  Burgett, Stanley G.
;    APPLICANT:  Kuhstoss, Stuart A.
;    APPLICANT:  Rao, Nagaraja R.
;    APPLICANT:  Richardson, Mark A.
;    APPLICANT:  Rosteck, Paul R., Jr.
;    TITLE OF INVENTION:  PLATENOLIDE SYNTHASE GENE
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  PAUL R. CANTRELL  1138
;      STREET:  LILLY CORPORATE CENTER
;      CITY:  INDIANAPOLIS
;      STATE:  IN
;      COUNTRY:  USA
;      ZIP:  46285
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  Macintosh
;      OPERATING SYSTEM:  Macintosh 7.0
;      SOFTWARE:  Microsoft Word 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/804,198
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  CANTRELL, PAUL R.
;      REGISTRATION NUMBER:  36,470
;      REFERENCE/DOCKET NUMBER:  P9113
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  317-276-3885
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  44377 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  350..14002
;    FEATURE:
;      NAME/KEY:  CDS
```

```
;       LOCATION:  14046..20036
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  20110..31284
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  31329..36071
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  36155..41830
US-08-804-198-1

  Query Match             56.8%;  Score 732.2;  DB 2;  Length 44377;
  Best Local Similarity   73.0%;  Pred. No. 4.7e-144;
  Matches  941;  Conservative    0;  Mismatches  348;  Indels    0;  Gaps    0;

Qy          1 GAGCCGATCGCCATCGTTGGCATGGCCTGTCGTTTCCCCGGCGGAGTGACCTCGGCGGAC 60
              |||||||||||| ||||| |||||||| || |||| ||| ||||| ||| | ||| |||||
Db      14148 GAGCCGATCGCGATCGTGGGCATGGCGTGCCGTTACCCGGGCGGTGTGGCGTCGCCGGAC 14207

Qy         61 GACTTCTGGGATCTGATCTCCTCCGAGCAGGACGCGATCGGCGGATTCCCCACCGACCGC 120
              ||| | ||||| ||| |   | |||     ||||||| ||   ||  ||||||  ||||||
Db      14208 GACCTGTGGGACCTGGTGGCAGCCGGTACGGACGCGGTCTCCGCGTTCCCCGTCGACCGT 14267

Qy        121 GGCTGGGACCTGGACACGCTCTACGACCCCGACCCCGACCACCCCGGCACCTGCTACACC 180
              |||||||| | ||   ||| |||||||||| || ||       ||     | |||
Db      14268 GGCTGGGACGTCGAGGGGCTGTACGACCCCGATCCGGAGGCGGTGGGGCGTAGTTACGTG 14327

Qy        181 CGAAACGGCGGATTCCTCTACGACGCAGGCCACTTCGACGCCGAATTCTTCGGCATCAGC 240
              || ·| ||||| |||||  ||   || | | ||||||||| || |||||||| |||
Db      14328 CGGGAGGGCGGGTTCCTGCACTCGGCGGCCGAGTTCGACGCGGAGTTCTTCGGGATCTCG 14387

Qy        241 CCCCGCGAAGCCCTCGCCATGGACCCCCAGCAACGACTCCTCCTCGAAACCGCCTGGGAA 300
              ||||| || ||    || ||||| || ||||| ||  | || || || ||  | |||||
Db      14388 CCCCGTGAGGCGGCGGCGATGGATCCGCAGCAGCGGTTGCTGCTGGAGACGTCGTGGGAG 14447

Qy        301 ACCATCGAACACGCCGGCATCAACCCCCACACCCTCCACGGCACCCCCCACCGGAGTCTTC 360
              |  | || |  || || |||  ||||    | || | ||||||||  ||||| ||||||
Db      14448 GCGCTGGAGCGGGCGGGGATCGTCCCCGCGTCGCTGCGCGGCACCCGTACCGGCGTCTTC 14507

Qy        361 ACCGGCACCAACGGACAGGACTACGCACTTCGCGTGCACAACGCGGGCCAGTCAACCGAT 420
              ||||||  ||      | |||||||     ||  |  ||   ||     | |     |||
Db      14508 ACCGGCGTCATGTACGACGACTACGGGTCGCGGTTCGACTCGGCTCCGCCGGAGTACGAG 14567

Qy        421 GGTTTCGCACTGACCGGAACCGCCGGCAGCGTCATCTCCGGTCGTATCTCGTACACGTTT 480
              || | |    ||| ||| | |||||||||| ||    |||||||  |  | ||  ||||
Db      14568 GGCTACCTCGTGAACGGCAGCGCCGGCAGCATCGCGTCCGGTCGGGTTGCCTATGCGTTG 14627

Qy        481 GGTTTTGAGGGTCCTGCGGTGTCGGTGGACACGGCTTGTTCCTCGTCGTTGGTGGCTTTG 540
              || || ||||| || ||| || |||||||||||||| ||||| ||||||||||||| |||
Db      14628 GGGTTGGAGGGGCCGGCGCTGACGGTGGACACGGCGTGTTCGTCGTCGTTGGTGGCGTTG 14687

Qy        541 CATCTGGCCTGTCAGGCGTTGCGTGCGGGTGAGTGCTCGATGGCGCTTGCCGGGGGTGTG 600
              ||||||||    ||| |||||||   |||||||||      ||||| | ||||| || |||
Db      14688 CATCTGGCGGTGCAGTCGTTGCGGCGGGGTGAGTGTGATCTGGCGTTGGCCGGTGGGGTG 14747

Qy        601 ACGGTGATGTCGTCTCCGGGTGCCTTCGTGGAGTTTTCGCGGCAGCGGGGTCTGGCCGCG 660
              ||||||||| || | |||    |   |||||||||| |||||||||||||| ||||| |||
Db      14748 ACGGTGATGGCGACGCCGACGGTGCTCGTGGAGTTCTCGCGGCAGCGGGGGCTGGCGGCG 14807

Qy        661 GACGGGCATTGCAAGGCGTTCTCGGCGGCGGCGGACGGGACCGGCTGGGGTGAGGGTGTG 720
              |||||||  |||||||||||| ||| ||  |||||||||| |  ||||  ||||||||||
Db      14808 GACGGGCGGTGCAAGGCGTTCGCGGAGGGTGCGGACGGGACGGCGTGGGCCGAGGGTGTG 14867

Qy        721 GGGATGCTGCTGGTGGAGCGGCTCTCCGACGCCCATCGCAACGGTCACCGTGTCCTGGCC 780
              ||  |||||||||||||||||||||||||||||||  ||||| || || || || |||||
Db      14868 GGCGTGCTGCTGGTGGAGCGGCTCTCCGACGCCCGCCGCAATGGCCATCGGGTGCTGGCG 14927

Qy        781 GTGGTGCGTGGCAGTGCGGTCAACCAGGACGGTGCGAGCAACGGTCTGACCGCGCCCAAC 840
              |||||||| |||||||||||||| |||||||||||||||||||| ||||| ||||| |
Db      14928 GTGGTGCGGGGCAGTGCGGTCAATCAGGACGGTGCGAGCAACGGGCTGACGGCGCCGAGT 14987

Qy        841 GGGCCGTCCCAGCAGCGTGTCATCCGCCAGGCCCTCGCCAACGCCGGCTTGTCGGCCGGT 900
              || ||  | |||||||| || |||||  |||| || ||| |||| ||  || || |||
Db      14988 GGTCCTGCGCAGCAGCGGGTGATCCGTGAGGCGCTGGCCGACGCGGGGCTGACGCCCGCC 15047

Qy        901 GATGTCGACGCGGTGGAGGCCCACGGCACCGGCACCACTTTGGGCGACCCGATCGAGGCC 960
              || |||||||||| ||||| |||||||||||||  |  |||||||||| ||||||||
Db      15048 GACGTCGACGCGGTCGAGGCGCACGGCACCGGCACACCCCTGGGCGACCCCATCGAGGCG 15107

Qy        961 CAGGCCCTCCTCGCGACCTACGGACAGGACCGTGCCGGCGAGGGGCCGCTGTGGCTGGGC 1020
                 ||  | || || ||||| ||     || ||     ||| | || ||| ||||| ||||
Db      15108 GGTGCGTTGCTGGCCACCTATGGCAGTGAGCGCCAGGGCCAAGGTCCGTTGTGGTTGGGG 15167

Qy       1021 TCGGTCAAGTCCAATGTCGGTCACACACAGGCTGCCGCGGGCGTCGCCGGGGTGATCAAG 1080
              ||| | ||||| ||   |||| ||  | ||||| || ||||| || |  || |||||||||
Db      15168 TCGTTGAAGTCGAACATCGGGCATGCGCAGGCGGCTGCGGGTGTGGGTGGCGTGATCAAG 15227

Qy       1081 ATGGTGATGGCGCTGCGGCATGGTCTGCTGCCGCGGACGTTGCATGTGGATGAGCCGTCG 1140
              |||||  |||| ||||||||||   | |||||||||| |||||||||| ||||||| |||
Db      15228 GTGGTGCAGGCGATGCGGCATGGGTCGTTGCCGCGGACGCTGCATGTGGATGCGCCGTCG 15287

Qy       1141 CCGCATGTGGACTGGTCCGCGGGTGCGGTGCAGCTGCTGACGGAGACGGTGCCCTGGCCC 1200
              || | ||||| ||| |  |||||||||||| |||||||||| |||||    | | |||||
Db      15288 TCGAAGGTGGAGTGGGCTTCGGGTGCGGTGGAGCTGCTGACCGAGACCCGGTCGTGGCCG 15347

Qy       1201 GGCGGGGAGGGGCGGCTACGGCGGGCAGGAGTGTCATCATTCGGCGTCAGCGGCACCAAC 1260
              |  ||| || |||| | |||||||| |  ||||| | |||| || ||||| ||||||
Db      15348 CGGCGGGTGGAGCGGGTGCGGCGGGCCGCGGTGTCGGCGTTCGGGGTGAGCGGGACCAAC 15407

Qy       1261 GCCCACGTCATCCTCGAAGAAGCACCCGC 1289
              ||||| ||  |||| || ||||| || ||
Db      15408 GCCCATGTGGTCCTGGAGGAAGCGCCGGC 15436


RESULT 3
US-08-804-227C-13
; Sequence 13, Application US/08804227C
; Patent No. 5876991
;  GENERAL INFORMATION:
;    APPLICANT:  DeHoff, Bradley S.
;    APPLICANT:  Kuhstoss, Stuart A.
;    APPLICANT:  Rosteck, Paul R., Jr.
;    APPLICANT:  Sutton, Kimberly L.
;    TITLE OF INVENTION:  POLYKETIDE SYNTHASE GENES
;    NUMBER OF SEQUENCES:  15
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  THOMAS G. PLANT  1501
;      STREET:  LILLY CORPORATE CENTER
;      CITY:  INDIANAPOLIS
;      STATE:  IN
;      COUNTRY:  USA
;      ZIP:  46285
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  MS-DOS
;      SOFTWARE:  ASCI(DOS) Text only
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/804,227C
;      FILING DATE:  February 21, 1997
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Plant, Thomas, G.
;      REGISTRATION NUMBER:  35,784
;      REFERENCE/DOCKET NUMBER:  X-8231
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  317-276-2459
;  INFORMATION FOR SEQ ID NO:  13:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  13987 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
```

```
;       NAME/KEY:  CDS
;       LOCATION:  350..13987
US-08-804-227C-13

  Query Match             56.0%;  Score 722.6;  DB 2;  Length 13987;
  Best Local Similarity   73.1%;  Pred. No. 3.8e-142;
  Matches  942;  Conservative    0;  Mismatches  344;  Indels    3;  Gaps    1;

Qy          1 GAGCCGATCGCCATCGTTGGCATGGCCTGTCGTTTCCCCGGCGGAGTGACCTCGGCGGAC 60
              |||||||||||| ||||| |||||||| || |||| ||| ||||| ||| | ||| |||||
Db       8255 GAGCCGATCGCGATCGTGGGCATGGCGTGCCGTTACCCGGGCGGTGTGGCGTCGCCGGAC 8314

Qy         61 GACTTCTGGGATCTGATCTCCTCCGAGCAGGACGCGATCGGCGGATTCCCCACCGACCGC 120
              ||| | ||||| ||| |  ||   ||   | || || ||  |   |||||  |||||||
Db       8315 GACCTGTGGGACCTGGTGGCCGGTGACGGGCACACGCTCTCCCCGTTCCCGGCCGACCGT 8374

Qy        121 GGCTGGGACCTGGACACGCTCTACGACCCCGACCCCGACCACCCCGGCACCTGCTACACC 180
              |||||||| | ||   ||| |||||||| || || |      || ||||   ||||
Db       8375 GGCTGGGACGTCGAGGGGCTGTACGACCCGGAGCCGGGGGTGCCGGGCAAGAGCTATGTA 8434

Qy        181 CGAAACGGCGGATTCCTCTACGACGCAGGCCACTTCGACGCCGAATTCTTCGGCATCAGC 240
              ||  | ||||| |||||       ||| | | | |||||||| || |||||||| ||
Db       8435 CGGGAAGGCGGGTTCCTGCGTTCCGCGGCCGAGTTCGACGCGGAGTTCTTCGGGATATCG 8494

Qy        241 CCCCGCGAAGCCCTCGCCATGGACCCCCAGCAACGACTCCTCCTCGAAACCGCCTGGGAA 300
              || ||||| |||   |||||||||||| ||||| ||  | || || ||  | |||||
Db       8495 CCGCGCGAGGCCACGGCCATGGACCCGCAGCAGCGGTTGCTGCTGGAGACGTCGTGGGAG 8554

Qy        301 ACCATCGAACACGCCGGCATCAACCCCCACACCCTCCACGGCACCCCCACCGGAGTCTTC 360
              |  | || |  |||||||||    ||  || | || | ||||||||   ||||| ||||||
Db       8555 GCGCTGGAGCGGGCCGGCATCGTTCCGGACTCGCTGCGCGGCACCCGGACCGGTGTCTTC 8614

Qy        361 ACCGGCACCAACGGACAGGACTACGCACTTCGCGTGCACAACGCGGGCCAGTCAACCGAT 420
              | ||||| |  |   ||||||||||||     |   ||    |||| | | |  |   ||
Db       8615 AGCGGCATCTCCCAGCAGGACTACGCGACCCAGCTGGGGGACGCCGCCGACACCTACGGC 8674

Qy        421 GGTTTCGCACTGACCGGAACCGCCGGCAGCGTCATCTCCGGTCGTATCTCGTACACGTTT 480
              ||    |  || || ||  |||||| || ||||||||||  |   | ||  ||||
Db       8675 GGGCATGTGCTCACGGGGACCCTCGGCAGTGTGATCTCCGGTCGGGTTGCCTATGCGTTG 8734

Qy        481 GGTTTTGAGGGTCCTGCGGTGTCGGTGGACACGGCTTGTTCCTCGTCGTTGGTGGCTTTG 540
              || || ||||| || ||| || |||||||||||||| ||||| ||||||||||||||| |||
Db       8735 GGGTTGGAGGGGCCGGCGCTGACGGTGGACACGGCGTGTTCGTCGTCGTTGGTGGCGTTG 8794

Qy        541 CATCTGGCCTGTCAGGCGTTGCGTGCGGGTGAGTGCTCGATGGCGCTTGCCGGGGGTGTG 600
              ||||||||    ||| |||||||   |||||||||     ||||| | ||||| || |||
Db       8795 CATCTGGCGGTGCAGTCGTTGCGGCGGGGTGAGTGTGATCTGGCGTTGGCCGGTGGGGTG 8854

Qy        601 ACGGTGATGTCGTCTCCGGGTGCCTTCGTGGAGTTTTCGCGGCAGCGGGGTCTGGCCGCG 660
              |||||||| || | |||   |  |||||||||| |||||||||||||| ||||| .|||
Db       8855 ACGGTGATGGCGACGCCGACGGTGTTCGTGGAGTTCTCGCGGCAGCGGGGGCTGGCGGCG 8914

Qy        661 GACGGGCATTGCAAGGCGTTCTCGGCGGCGGCGGACGGGACCGGCTGGGGTGAGGGTGTG 720
              |||||||  |||||||||||| ||| ||  ||||||||||| |  ||||  |||||||||
Db       8915 GACGGGCGGTGCAAGGCGTTCGCGGAGGGTGCGGACGGGACGGCGTGGGCGGAGGGTGTG 8974

Qy        721 GGGATGCTGCTGGTGGAGCGGCTCTCCGACGCCCATCGCAACGGTCACCGTGTCCTGGCC 780
              ||  |||||||||||||||||| ||||||||  | ||||||||||| || || |||||
Db       8975 GGTGTGCTGCTGGTGGAGCGGCTTTCCGACGCGCGCCGCAACGGTCATCGGGTGCTGGCG 9034

Qy        781 GTGGTGCGTGGCAGTGCGGTCAACCAGGACGGTGCGAGCAACGGTCTGACCGCGCCCAAC 840
              |||||||| |||||||||||||||| |||||||||||||||| || ||||| ||||| |
Db       9035 GTGGTGCGGGGCAGTGCGGTCAATCAGGACGGTGCGAGCAATGGGCTGACGGCGCCGAGT 9094

Qy        841 GGGCCGTCCCAGCAGCGTGTCATCCGCCAGGCCCTCGCCAACGCCGGCTTGTCGGCCGGT 900
              || ||| | |||||||| || |||||  |||| || ||  | || ||  ||  | |||
Db       9095 GGTCCGGCGCAGCAGCGGGTGATCCGTGAGGCGCTGGCTGATGCGGGGCTGGTGCCCGCC 9154

Qy        901 GATGTCGACGCGGTGGAGGCCCACGGCACCGGCACCACTTTGGGCGACCCGATCGAGGCC 960
              || || || | ||||||||| ||||| || || ||  |  |||| || ||||||||||
Db       9155 GACGTGGATGTGGTGGAGGCGCACGGTACGGGGACGGCGCTGGGTGATCCGATCGAGGCG 9214

Qy        961 CAGGCCCTCCTCGCGACCTACGGACAGGACCGTGCCGGCGAGGGGCCGCTGTGGCTGGGC 1020
```

```
                ||| || || || || ||||| | ||| || | ||||||    ||| ||||||| ||
Db       9215 GGTGCGCTGCTGGCCACGTACGGGCGGGAGCGGGTCGGCGA---TCCGTTGTGGCTCGGG 9271

Qy       1021 TCGGTCAAGTCCAATGTCGGTCACACACAGGCTGCCGCGGGCGTCGCCGGGGTGATCAAG 1080
              ||| | ||||| ||  |||| ||  | ||||| || ||||| || |  || |||||||||
Db       9272 TCGTTGAAGTCGAACATCGGGCATGCGCAGGCGGCTGCGGGTGTGGGTGGTGTGATCAAG 9331

Qy       1081 ATGGTGATGGCGCTGCGGCATGGTCTGCTGCCGCGGACGTTGCATGTGGATGAGCCGTCG 1140
              |||||  || | |||||||||  | ||||||||||| |||||||||| |||||||
Db       9332 GTGGTGCAGGGGATGCGGCATGGGTCGTTGCCGCGGACGCTGCATGTGGATGCGCCGTCG 9391

Qy       1141 CCGCATGTGGACTGGTCCGCGGGTGCGGTGCAGCTGCTGACGGAGACGGTGCCCTGGCCC 1200
              || | ||||| ||| |  |||||||||| |||||||||| |||||   | | |||||
Db       9392 TCGAAGGTGGAGTGGGCTTCGGGTGCGGTGGAGCTGCTGACCGAGACCCGGTCGTGGCCG 9451

Qy       1201 GGCGGGGAGGGGCGGCTACGGCGGGCAGGAGTGTCATCATTCGGCGTCAGCGGCACCAAC 1260
              |  ||| || |||| | |||||||| |  ||||| | ||||| || |||| ||||||
Db       9452 CGGCGGGTGGAGCGGGTGCGGCGGGCCGCGGTGTCGGCGTTCGGGGTGAGCGGGACCAAC 9511

Qy       1261 GCCCACGTCATCCTCGAAGAAGCACCCGC 1289
              ||||| ||  |||| || ||||| || ||
Db       9512 GCCCATGTGGTCCTGGAGGAAGCGCCGGC 9540


RESULT 4
US-09-105-537-32
; Sequence 32, Application US/09105537A
; Patent No. 6265202
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT: Liu, H.
;  APPLICANT: Xue, Y.
;  APPLICANT: Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/105,537A
;  CURRENT FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: PastSEQ for Windows Version 3.0
; SEQ ID NO 32
;   LENGTH: 11220
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-105-537-32

  Query Match             55.5%;  Score 715.4;  DB 4;  Length 11220;
  Best Local Similarity   72.6%;  Pred. No. 1.2e-140;
  Matches  939;  Conservative    0;  Mismatches  351;  Indels    3;  Gaps    1;

Qy          1 GAGCCGATCGCCATCGTTGGCATGGCCTGTCGTTTCCCCGGCGGAGTGACCTCGGCGGAC 60
              |||||  |||| ||||| |||||||| || ||  | || ||||| ||  |||||  ||||
Db       4618 GAGCCCGTCGCGATCGTCGGCATGGCGTGCCGCCTGCCGGGCGGTGTCGCCTCGCCGGAG 4677

Qy         61 GACTTCTGGGATCTGATCTCCTCCGAGCAGGACGCGATCGGCGGATTCCCCACCGACCGC 120
              ||| | |||   ||| |  ||  ||   ||||||||||  ||| |||||    ||||||
Db       4678 GACCTGTGGCGGCTGGTGGCCGGCGGCGAGGACGCGATCTCCGGCTTCCCGCAGGACCGC 4737

Qy        121 GGCTGGGACCTGGACACGCTCTACGACCCCGACCCCGACCACCCCGGCACCTGCTACACC 180
              |||||||| ||||    ||| |||||||| ||||| |||     ||||         ||| |
Db       4738 GGCTGGGACGTGGAGGGGCTGTACGACCCGGACCCGGACGCGTCCGGGCGGACGTACTGC 4797

Qy        181 CGAAACGGCGGATTCCTCTACGACGCAGGCCACTTCGACGCCGAATTCTTCGGCATCAGC 240
              ||   ||| || |||||| |||| || ||| | |||||||||| ||||||| |||
Db       4798 CGTGCCGGTGGCTTCCTCGACGAGGCGGGCGAGTTCGACGCCGACTTCTTCGGGATCTCG 4857

Qy        241 CCCCGCGAAGCCCTCGCCATGGACCCCCAGCAACGACTCCTCCTCGAAACCGCCTGGGAA 300
              || ||||| |||||||||||||||||| ||||| || |||||||| || ||| |||||||
Db       4858 CCGCGCGAGGCCCTCGCCATGGACCCGCAGCAGCGGCTCCTCCTGGAGACCTCCTGGGAG 4917

Qy        301 ACCATCGAACACGCCGGCATCAACCCCCACACCCTCCACGGCACCCCCACCGGAGTCTTC 360
              || ||||  ||||||| ||| ||||    | |||| || ||     |    ||| || |||
Db       4918 GCCGTCGAGGACGCCGGGATCGACCCGACCTCCCTTCAGGGGCAGCAGGTCGGCGTGTTC 4977
```

```
Qy        361 ACCGGCACCAACGGACAGGACTACGCACTTCGCGTGCACAACGCGGGCCAGTCAACCGAT 420
Db       4978 GCGGGCACCAACGGCCCCCACTACGAGCCGCTGCTCCGCAACACCGCCGAGGATCTTGAG 5037

Qy        421 GGTTTCGCACTGACCGGAACCGCCGGCAGCGTCATCTCCGGTCGTATCTCGTACACGTTT 480
Db       5038 GGTTACGTCGGGACGGGCAACGCCGCCAGCATCATGTCGGGCCGTGTCTCGTACACCCTC 5097

Qy        481 GGTTTTGAGGGTCCTGCGGTGTCGGTGGACACGGCTTGTTCCTCGTCGTTGGTGGCTTTG 540
Db       5098 GGCCTGGAGGGCCCGGCCGTCACGGTCGACACCGCCTGCTCCTCCTCGCTGGTCGCCCTG 5157

Qy        541 CATCTGGCCTGTCAGGCGTTGCGTGCGGGTGAGTGCTCGATGGCGCTTGCCGGGGGTGTG 600
Db       5158 CACCTCGCCGTGCAGGCCCTGCGCAAGGGCGAATGCGGACTGGCGCTCGCGGGCGGTGTG 5217

Qy        601 ACGGTGATGTCGTCTCCGGGTGCCTTCGTGGAGTTTTCGCGGCAGCGGGGTCTGGCCGCG 660
Db       5218 ACGGTCATGTCGACGCCCACGACGTTCGTGGAGTTCAGCCGGCAGCGCGGGCTCGCGGAG 5277

Qy        661 GACGGGCATTGCAAGGCGTTCTCGGCGGCGGCGGACGGGACCGGCTGGGGTGAGGGTGTG 720
Db       5278 GACGGCCGGTCGAAGGCGTTCGCCGCGTCGGCGGACGGCTTCGGCCCGGCGGAGGGCGTC 5337

Qy        721 GGGATGCTGCTGGTGGAGCGGCTCTCCGACGCCCATCGCAACGGTCACCGTGTCCTGGCC 780
Db       5338 GGCATGCTCCTCGTCGAGCGCCTGTCGGACGCCCGCCGCAACGGACACCGTGTGCTGGCG 5397

Qy        781 GTGGTGCGTGGCAGTGCGGTCAACCAGGACGGTGCGAGCAACGGTCTGACCGCGCCCAAC 840
Db       5398 GTCGTGCGCGGCAGCGCGGTCAACCAGGACGGCGCGAGCAACGGCCTGACCGCCCCGAAC 5457

Qy        841 GGGCCGTCCCAGCAGCGTGTCATCCGCCAGGCCCTCGCCAACGCCGGCTTGTCGGCCGGT 900
Db       5458 GGGCCCTCGCAGCAGCGCGTCATCCGGCGCGCGCTCGCGGACGCCCGACTGACGACCGCC 5517

Qy        901 GATGTCGACGCGGTGGAGGCCCACGGCACCGGCACCACTTTGGGCGACCCGATCGAGGCC 960
Db       5518 GACGTGGACGTCGTCGAGGCCCACGGCACGGGCACGCGACTCGGCGACCCGATCGAGGCA 5577

Qy        961 CAGGCCCTCCTCGCGACCTACGGACAGGACCGTGCCGGCGAGGGGCCGCTGTGGCTGGGC 1020
Db       5578 CAGGCCCTCATCGCCACCTACGGCCAGGGGCGCGACACCGAACAGCCGCTGCGCCTGGGG 5637

Qy       1021 TCGGTCAAGTCCAATGTCGGTCACACACAGGCTGCCGCGGGCGTCGCCGGGGTGATCAAG 1080
Db       5638 TCGTTGAAGTCCAACATCGGACACACCCAGGCCGCCGCCGGTGTCTCCGGCATCATCAAG 5697

Qy       1081 ATGGTGATGGCGCTGCGGCATGGTCTGCTGCCGCGGACGTTGCATGTGGATGAGCCGTCG 1140
Db       5698 ATGGTCCAGGCGATGCGCCACGGCGTCCTGCCGAAGACGCTCCACGTGGACCGGCCGTCG 5757

Qy       1141 CCGCATGTGGACTGGTCCGCGGGTGCGGTGCAGCTGCTGACGGAGACGGTGCCCTGGCCC 1200
Db       5758 GACCAGATCGACTGGTCGGCGGGCACGGTCGAGCTGCTCACCGAGGCCATGGACTGGCCG 5817

Qy       1201 GG---CGGGGAGGGGCGGCTACGGCGGGCAGGAGTGTCATCATTCGGCGTCAGCGGCACC 1257
Db       5818 AGGAAGCAGGAGGGCGGGCTGCGCCGCGCGGCCGTCTCCTCCTTCGGCATCAGCGGCACG 5877

Qy       1258 AACGCCCACGTCATCCTCGAAGAAGCACCCGCC 1290
Db       5878 AACGCGCACATCGTGCTCGAAGAAGCCCCGGTC 5910


RESULT 5
US-09-105-537-5
; Sequence 5, Application US/09105537A
; Patent No. 6265202
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT: Liu, H.
;  APPLICANT: Xue, Y.
;  APPLICANT: Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/105,537A
;  CURRENT FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 5
;   LENGTH: 36778
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-105-537-5

  Query Match             55.5%;  Score 715.4;  DB 4;  Length 36778;
  Best Local Similarity   72.6%;  Pred. No. 1.4e-140;
  Matches  939;  Conservative    0;  Mismatches  351;  Indels    3;  Gaps    1;

Qy          1 GAGCCGATCGCCATCGTTGGCATGGCCTGTCGTTTCCCCGGCGGAGTGACCTCGGCGGAC 60
Db      20305 GAGCCCGTCGCGATCGTCGGCATGGCGTGCCGCCTGCCGGGCGGTGTCGCCTCGCCGGAG 20364

Qy         61 GACTTCTGGGATCTGATCTCCTCCGAGCAGGACGCGATCGGCGGATTCCCCACCGACCGC 120
Db      20365 GACCTGTGGCGGCTGGTGGCCGGCGGCGAGGACGCGATCTCCGGCTTCCCGCAGGACCGC 20424

Qy        121 GGCTGGGACCTGGACACGCTCTACGACCCCGACCCCGACCACCCCGGCACCTGCTACACC 180
Db      20425 GGCTGGGACGTGGAGGGGCTGTACGACCCGGACCCGGACGCGTCCGGGCGGACGTACTGC 20484

Qy        181 CGAAACGGCGGATTCCTCTACGACGCAGGCCACTTCGACGCCGAATTCTTCGGCATCAGC 240
Db      20485 CGTGCCGGTGGCTTCCTCGACGAGGCGGGCGAGTTCGACGCCGACTTCTTCGGGATCTCG 20544

Qy        241 CCCCGCGAAGCCCTCGCCATGGACCCCCAGCAACGACTCCTCCTCGAAACCGCCTGGGAA 300
Db      20545 CCGCGCGAGGCCCTCGCCATGGACCCGCAGCAGCGGCTCCTCCTGGAGACCTCCTGGGAG 20604

Qy        301 ACCATCGAACACGCCGGCATCAACCCCCACACCCTCCACGGCACCCCCACCGGAGTCTTC 360
Db      20605 GCCGTCGAGGACGCCGGGATCGACCCGACCTCCCTTCAGGGGCAGCAGGTCGGCGTGTTC 20664

Qy        361 ACCGGCACCAACGGACAGGACTACGCACTTCGCGTGCACAACGCGGGCCAGTCAACCGAT 420
Db      20665 GCGGGCACCAACGGCCCCCACTACGAGCCGCTGCTCCGCAACACCGCCGAGGATCTTGAG 20724

Qy        421 GGTTTCGCACTGACCGGAACCGCCGGCAGCGTCATCTCCGGTCGTATCTCGTACACGTTT 480
Db      20725 GGTTACGTCGGGACGGGCAACGCCGCCAGCATCATGTCGGGCCGTGTCTCGTACACCCTC 20784

Qy        481 GGTTTTGAGGGTCCTGCGGTGTCGGTGGACACGGCTTGTTCCTCGTCGTTGGTGGCTTTG 540
Db      20785 GGCCTGGAGGGCCCGGCCGTCACGGTCGACACCGCCTGCTCCTCCTCGCTGGTCGCCCTG 20844

Qy        541 CATCTGGCCTGTCAGGCGTTGCGTGCGGGTGAGTGCTCGATGGCGCTTGCCGGGGGTGTG 600
Db      20845 CACCTCGCCGTGCAGGCCCTGCGCAAGGGCGAATGCGGACTGGCGCTCGCGGGCGGTGTG 20904

Qy        601 ACGGTGATGTCGTCTCCGGGTGCCTTCGTGGAGTTTTCGCGGCAGCGGGGTCTGGCCGCG 660
Db      20905 ACGGTCATGTCGACGCCCACGACGTTCGTGGAGTTCAGCCGGCAGCGCGGGCTCGCGGAG 20964

Qy        661 GACGGGCATTGCAAGGCGTTCTCGGCGGCGGCGGACGGGACCGGCTGGGGTGAGGGTGTG 720
Db      20965 GACGGCCGGTCGAAGGCGTTCGCCGCGTCGGCGGACGGCTTCGGCCCGGCGGAGGGCGTC 21024

Qy        721 GGGATGCTGCTGGTGGAGCGGCTCTCCGACGCCCATCGCAACGGTCACCGTGTCCTGGCC 780
Db      21025 GGCATGCTCCTCGTCGAGCGCCTGTCGGACGCCCGCCGCAACGGACACCGTGTGCTGGCG 21084

Qy        781 GTGGTGCGTGGCAGTGCGGTCAACCAGGACGGTGCGAGCAACGGTCTGACCGCGCCCAAC 840
Db      21085 GTCGTGCGCGGCAGCGCGGTCAACCAGGACGGCGCGAGCAACGGCCTGACCGCCCCGAAC 21144
```

```
Qy        841 GGGCCGTCCCAGCAGCGTGTCATCCGCCAGGCCCTCGCCAACGCCGGCTTGTCGGCCGGT 900
              |||||| || |||||||| |||||||| |  || |||||  |||||  |  || || |||
Db      21145 GGGCCCTCGCAGCAGCGCGTCATCCGGCGCGCGCTCGCGGACGCCCGACTGACGACCGCC 21204

Qy        901 GATGTCGACGCGGTGGAGGCCCACGGCACCGGCACCACTTTGGGCGACCCGATCGAGGCC 960
              || || ||||  || ||||||||||||||| |||||      | ||||||||||||||||||
Db      21205 GACGTGGACGTCGTCGAGGCCCACGGCACGGGCACGCGACTCGGCGACCCGATCGAGGCA 21264

Qy        961 CAGGCCCTCCTCGCGACCTACGGACAGGACCGTGCCGGCGAGGGGCCGCTGTGGCTGGGC 1020
              ||||||||| |||| |||||||| ||||  || | |  |||   ||||||| | |||||
Db      21265 CAGGCCCTCATCGCCACCTACGGCCAGGGGCGCGACACCGAACAGCCGCTGCGCCTGGGG 21324

Qy       1021 TCGGTCAAGTCCAATGTCGGTCACACACAGGCTGCCGCGGGCGTCGCCGGGGTGATCAAG 1080
              ||| | ||||||||  |||| ||||| |||||  ||||| || ||| ||||  | ||||||
Db      21325 TCGTTGAAGTCCAACATCGGACACACCCAGGCCGCCGCCGGTGTCTCCGGCATCATCAAG 21384

Qy       1081 ATGGTGATGGCGCTGCGGCATGGTCTGCTGCCGCGGACGTTGCATGTGGATGAGCCGTCG 1140
              |||||   ||| |||| || ||  | ||||||  |||| | || |||||  |||||||
Db      21385 ATGGTCCAGGCGATGCGCCACGGCGTCCTGCCGAAGACGCTCCACGTGGACCGGCCGTCG 21444

Qy       1141 CCGCATGTGGACTGGTCCGCGGGTGCGGTGCAGCTGCTGACGGAGACGGTGCCCTGGCCC 1200
                ||  | |||||||| |||||  ||||  ||||||| || ||| |  ||  ||||||
Db      21445 GACCAGATCGACTGGTCGGCGGGCACGGTCGAGCTGCTCACCGAGGCCATGGACTGGCCG 21504

Qy       1201 GG---CGGGGAGGGGCGGCTACGGCGGGCAGGAGTGTCATCATTCGGCGTCAGCGGCACC 1257
               |      ||||||  |||| || || || |  || || || |||||| ||||||||||
Db      21505 AGGAAGCAGGAGGGCGGGCTGCGCCGCGCGGCCGTCTCCTCCTTCGGCATCAGCGGCACG 21564

Qy       1258 AACGCCCACGTCATCCTCGAAGAAGCACCCGCC 1290
              ||||| ||| || | |||||||||||| || | |
Db      21565 AACGCGCACATCGTGCTCGAAGAAGCCCCGGTC 21597


RESULT 6
US-09-320-878-19
; Sequence 19, Application US/09320878A
; Patent No. 6117659
; GENERAL INFORMATION:
;  APPLICANT: ASHLEY, Gary
;  APPLICANT: BETLACH, Melanie C.
;  APPLICANT: BETLACH, Mary C.
;  APPLICANT: McDANIEL, Robert
;  APPLICANT: TANG, Li
;  TITLE OF INVENTION: RECOMBINANT NARBONOLIDE POLYKETIDE SYNTHASE
;  FILE REFERENCE: 300622002120
;  CURRENT APPLICATION NUMBER: US/09/320,878A
;  CURRENT FILING DATE: 1999-05-27
;  EARLIER APPLICATION NUMBER: CIP OF 09/141,908
;  EARLIER FILING DATE: 1998-08-28
;  EARLIER APPLICATION NUMBER: CIP OF 09/073,538
;  EARLIER FILING DATE: 1998-05-06
;  EARLIER APPLICATION NUMBER: CIP OF 08/846,247
;  EARLIER FILING DATE: 1997-04-30
;  EARLIER APPLICATION NUMBER: 60/119,139
;  EARLIER FILING DATE: 1999-02-08
;  EARLIER APPLICATION NUMBER: 60/100,880
;  EARLIER FILING DATE: 1998-09-22
;  EARLIER APPLICATION NUMBER: 60/087,080
;  EARLIER FILING DATE: 1998-05-28
;  NUMBER OF SEQ ID NOS: 34
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 19
;   LENGTH: 38506
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-320-878-19

  Query Match           55.5%;  Score 715.4;  DB 3;  Length 38506;
  Best Local Similarity   72.6%;  Pred. No. 1.5e-140;
  Matches  939;  Conservative    0;  Mismatches  351;  Indels    3;  Gaps    1;

Qy          1 GAGCCGATCGCCATCGTTGGCATGGCCTGTCGTTTCCCCGGCGGAGTGACCTCGGCGGAC 60
              |||||  |||| ||||| |||||||| || ||  | || ||||| ||  ||||| ||||

Db      18447 GAGCCCGTCGCGATCGTCGGCATGGCGTGCCGCCTGCCGGGCGGTGTCGCCTCGCCGGAG 18506

Qy         61 GACTTCTGGGATCTGATCTCCTCCGAGCAGGACGCGATCGGCGGATTCCCCACCGACCGC 120
              ||| | |||   ||| |  ||  ||  |||||||||||  ||| |||||     ||||||
Db      18507 GACCTGTGGCGGCTGGTGGCCGGCGGCGAGGACGCGATCTCCGGCTTCCCGCAGGACCGC 18566

Qy        121 GGCTGGGACCTGGACACGCTCTACGACCCCGACCCCGACCACCCCGGCACCTGCTACACC 180
              |||||||| ||||   ||| |||||||| ||||| |||    ||||        |||  |
Db      18567 GGCTGGGACGTGGAGGGGCTGTACGACCCGGACCCGGACGCGTCCGGGCGGACGTACTGC 18626

Qy        181 CGAAACGGCGGATTCCTCTACGACGCAGGCCACTTCGACGCCGAATTCTTCGGCATCAGC 240
              ||   ||| || |||||| |||| || ||| | |||||||||| ||||||| |||
Db      18627 CGTGCCGGTGGCTTCCTCGACGAGGCGGGCGAGTTCGACGCCGACTTCTTCGGGATCTCG 18686

Qy        241 CCCCGCGAAGCCCTCGCCATGGACCCCCAGCAACGACTCCTCCTCGAAACCGCCTGGGAA 300
              || ||||| |||||||||||||||| ||||| || |||||||| || ||| |||||||
Db      18687 CCGCGCGAGGCCCTCGCCATGGACCCGCAGCAGCGGCTCCTCCTGGAGACCTCCTGGGAG 18746

Qy        301 ACCATCGAACACGCCGGCATCAACCCCCACACCCTCCACGGCACCCCCACCGGAGTCTTC 360
              || |||| ||||||| ||| ||||    | |||| || ||     |   ||| || |||
Db      18747 GCCGTCGAGGACGCCGGGATCGACCCGACCTCCCTTCAGGGGCAGCAGGTCGGCGTGTTC 18806

Qy        361 ACCGGCACCAACGGACAGGACTACGCACTTCGCGTGCACAACGCGGGCCAGTCAACCGAT 420
              | |||||||||||  |   ||||||  |  |   | | |||| | | | ||      ||
Db      18807 GCGGGCACCAACGGCCCCCACTACGAGCCGCTGCTCCGCAACACCGCCGAGGATCTTGAG 18866

Qy        421 GGTTTCGCACTGACCGGAACCGCCGGCAGCGTCATCTCCGGTCGTATCTCGTACACGTTT 480
              |||| ||    ||| || | ||||| |||| |||| || ||.||| ||||||||||  |
Db      18867 GGTTACGTCGGGACGGGCAACGCCGCCAGCATCATGTCGGGCCGTGTCTCGTACACCCTC 18926

Qy        481 GGTTTTGAGGGTCCTGCGGTGTCGGTGGACACGGCTTGTTCCTCGTCGTTGGTGGCTTTG 540
              ||  | ||||| || || ||  |||| |||| || || ||||| ||| |||| ||  ||
Db      18927 GGCCTGGAGGGCCCGGCCGTCACGGTCGACACCGCCTGCTCCTCCTCGCTGGTCGCCCTG 18986

Qy        541 CATCTGGCCTGTCAGGCGTTGCGTGCGGGTGAGTGCTCGATGGCGCTTGCCGGGGGTGTG 600
              || || |||   ||||| |||    ||| || |||    ||||||| || || ||||||
Db      18987 CACCTCGCCGTGCAGGCCCTGCGCAAGGGCGAATGCGGACTGGCGCTCGCGGGCGGTGTG 19046

Qy        601 ACGGTGATGTCGTCTCCGGGTGCCTTCGTGGAGTTTTCGCGGCAGCGGGGTCTGGCCGCG 660
              ||||| |||||| | ||     | ||||||||||    |||||||| || || || | |
Db      19047 ACGGTCATGTCGACGCCCACGACGTTCGTGGAGTTCAGCCGGCAGCGCGGGCTCGCGGAG 19106

Qy        661 GACGGGCATTGCAAGGCGTTCTCGGCGGCGGCGGACGGGACCGGCTGGGGTGAGGGTGTG 720
              ||||| |  | |||||||||| | ||| |||||||||    |||| ||  ||||| ||
Db      19107 GACGGCCGGTCGAAGGCGTTCGCCGCGTCGGCGGACGGCTTCGGCCCGGCGGAGGGCGTC 19166

Qy        721 GGGATGCTGCTGGTGGAGCGGCTCTCCGACGCCCATCGCAACGGTCACCGTGTCCTGGCC 780
              || ||||| || || ||||| || || |||||||  |||||||| ||||||||  |||||
Db      19167 GGCATGCTCCTCGTCGAGCGCCTGTCGGACGCCCGCCGCAACGGACACCGTGTGCTGGCG 19226

Qy        781 GTGGTGCGTGGCAGTGCGGTCAACCAGGACGGTGCGAGCAACGGTCTGACCGCGCCCAAC 840
              || ||||| ||||| ||||||||||||||||| |||||||||| |||||||| || |||
Db      19227 GTCGTGCGCGGCAGCGCGGTCAACCAGGACGGCGCGAGCAACGGCCTGACCGCCCCGAAC 19286

Qy        841 GGGCCGTCCCAGCAGCGTGTCATCCGCCAGGCCCTCGCCAACGCCGGCTTGTCGGCCGGT 900
              ||||| || |||||||| |||||||  || ||||  |||||  |  || || |||
Db      19287 GGGCCCTCGCAGCAGCGCGTCATCCGGCGCGCGCTCGCGGACGCCCGACTGACGACCGCC 19346

Qy        901 GATGTCGACGCGGTGGAGGCCCACGGCACCGGCACCACTTTGGGCGACCCGATCGAGGCC 960
              || || ||||  || ||||||||||||||| |||||      | ||||||||||||||||||
Db      19347 GACGTGGACGTCGTCGAGGCCCACGGCACGGGCACGCGACTCGGCGACCCGATCGAGGCA 19406

Qy        961 CAGGCCCTCCTCGCGACCTACGGACAGGACCGTGCCGGCGAGGGGCCGCTGTGGCTGGGC 1020
              |||||||| |||| |||||||| ||||  || | |  |||   ||||||| | |||||
Db      19407 CAGGCCCTCATCGCCACCTACGGCCAGGGGCGCGACACCGAACAGCCGCTGCGCCTGGGG 19466

Qy       1021 TCGGTCAAGTCCAATGTCGGTCACACACAGGCTGCCGCGGGCGTCGCCGGGGTGATCAAG 1080
              ||| | ||||||||  |||| ||||| |||||  ||||| || ||| ||||  | ||||||
Db      19467 TCGTTGAAGTCCAACATCGGACACACCCAGGCCGCCGCCGGTGTCTCCGGCATCATCAAG 19526

Qy       1081 ATGGTGATGGCGCTGCGGCATGGTCTGCTGCCGCGGACGTTGCATGTGGATGAGCCGTCG 1140
              |||||   ||| |||| || ||  | ||||||  |||| | || |||||  |||||||
Db      19527 ATGGTCCAGGCGATGCGCCACGGCGTCCTGCCGAAGACGCTCCACGTGGACCGGCCGTCG 19586
```

```
Qy       1141 CCGCATGTGGACTGGTCCGCGGGTGCGGTGCAGCTGCTGACGGAGACGGTGCCCTGGCCC 1200
Db      19587 GACCAGATCGACTGGTCGGCGGGCACGGTCGAGCTGCTCACCGAGGCCATGGACTGGCCG 19646

Qy       1201 GG---CGGGGAGGGGCGGCTACGGCGGGCAGGAGTGTCATCATTCGGCGTCAGCGGCACC 1257
Db      19647 AGGAAGCAGGAGGGCGGGCTGCGCCGCGCGGCCGTCTCCTCCTTCGGCATCAGCGGCACG 19706

Qy       1258 AACGCCCACGTCATCCTCGAAGAAGCACCCGCC 1290
Db      19707 AACGCGCACATCGTGCTCGAAGAAGCCCCGGTC 19739


RESULT 7
US-09-036-987A-1
; Sequence 1, Application US/09036987A
; Patent No. 6143526
;  GENERAL INFORMATION:
;    APPLICANT:  Baltz, Richard H.
;    APPLICANT:  Broughton, Mary C.
;    APPLICANT:  Crawford, Kathryn P.
;    APPLICANT:  Madduri, Krishnamurthy
;    APPLICANT:  Merlo, Donald J.
;    APPLICANT:  Treadway, Patti J.
;    APPLICANT:  Turner, Jan R.
;    APPLICANT:  Waldron, Clive
;    TITLE OF INVENTION:  Biosynthetic Genes For Spinosyn Insecticide
;    TITLE OF INVENTION:  Production
;    NUMBER OF SEQUENCES:  39
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Dow AgroSciences LLC Patent Department
;      STREET:  9330 Zionsville Road
;      CITY:  Indianapolis
;      STATE:  Indiana
;      COUNTRY:  USA
;      ZIP:  46268
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/036,987A
;      FILING DATE:  09-MAR-1998
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Stuart, Donald R
;      REGISTRATION NUMBER:  28,479
;      REFERENCE/DOCKET NUMBER:  50,608
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (317)337-4816
;      TELEFAX:  (317)337-4847
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  80161 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-09-036-987A-1

  Query Match             55.4%;  Score 715.2;  DB 3;  Length 80161;
  Best Local Similarity   72.9%;  Pred. No. 1.8e-140;
  Matches  937;  Conservative    0;  Mismatches  343;  Indels    6;  Gaps    1;

Qy          1 GAGCCGATCGCCATCGTTGGCATGGCCTGTCGTTTCCCCGGCGGAGTGACCTCGGCGGAC 60
Db      35518 GAACCCATCGCGATCGTGGCGATGGGCTGCCGGTACCCGGGTGGGGTGCAGGACCCGGAA 35577

Qy         61 GACTTCTGGGATCTGATCTCCTCCGAGCAGGACGCGATCGGCGGATTCCCCACCGACCGC 120
Db      35578 GGGCTGTGGAAACTGGTCGCCTCCGGTGGCGACGCCATCGGTGAATTCCCCGCTGATCGT 35637

Qy        121 GGCTGGGACCTGGACACGCTCTACGACCCCGACCCCGACCACCCCGGCACCTGCTACACC 180
Db      35638 GGTTGGCACCTCGACGAGCTCTACGATCCCGACCCGGATCAGCCCGGAACCTGCTACACC 35697

Qy        181 CGAAACGGCGGATTCCTCTACGACGCAGGCCACTTCGACGCCGAATTCTTCGGCATCAGC 240
Db      35698 CGGCACGGCGGCTTCCTCCACGACGCCGGCGAGTTCGACGCGGGATTCTTCGACATCAGC 35757

Qy        241 CCCCGCGAAGCCCTCGCCATGGACCCCCAGCAACGACTCCTCCTCGAAACCGCCTGGGAA 300
Db      35758 CCCCGTGAGGCGCTCGCGATGGACCCGCAGCAGCGGCTGCTGCTGGAAATCTCCTGGGAG 35817

Qy        301 ACCATCGAACACGCCGGCATCAACCCCCACACCCTCCACGGCACCCCCACCGGAGTCTTC 360
Db      35818 ACCGTCGAATCCGCTGGGATGGACCCGAGGTCCTTGCGGGGGAGCCGCACCGGGGTGTTC 35877

Qy        361 ACCGGCACCAACGGACAGGACTACGCACTTCGCGTGCACAACGCGGGCCAGTCAACCGAT 420
Db      35878 GCGGGATTGATGTACGAGGGCTATGACACCGGCGCCCACCGGGCAGGAGAAGGTGTCGAA 35937

Qy        421 GGTTTCGCACTGACCGGAACCGCCGGCAGCGTCATCTCCGGTCGTATCTCGTACACGTTT 480
Db      35938 GGCTATCTCGGAACCGGCAATGCGGGAAGCGTCGCCTCTGGTCGGGTTGCGTATGCGTTC 35997

Qy        481 GGTTTTGAGGGTCCTGCGGTGTCGGTGGACACGGCTTGTTCCTCGTCGTTGGTGGCTTTG 540
Db      35998 GGGTTCGAGGGCCCAGCGGTGACGGTAGACACGGCGTGCTCGTCGTCGTTGGTGGCGCTG 36057

Qy        541 CATCTGGCCTGTCAGGCGTTGCGTGCGGGTGAGTGCTCGATGGCGCTTGCCGGGGGTGTG 600
Db      36058 CATTTGGCGTGTCAGTCGTTGCGGCAGGGCGAGTGTGATCTGGCGCTGGCCGGTGGAGTG 36117

Qy        601 ACGGTGATGTCGTCTCCGGGTGCCTTCGTGGAGTTTTCGCGGCAGCGGGGTCTGGCCGCG 660
Db      36118 ACGGTGATGTCGACGCCGGAGAGGTTCGTGGAGTTCTCCCGTCAGCGTGGTCTCGCACCG 36177

Qy        661 GACGGGCATTGCAAGGCGTTCTCGGCGGCGGCGGACGGGACCGGCTGGGGTGAGGGTGTG 720
Db      36178 GATGGGCGGTGTAAGTCGTTCGCGGCGGCTGCGGATGGAACCGGTTGGGGTGAGGGTGCC 36237

Qy        721 GGGATGCTGCTGGTGGAGCGGCTCTCCGACGCCCATCGCAACGGTCACCGTGTCCTGGCC 780
Db      36238 GGTTTGGTGTTGCTGGAGCGGCTGTCAGACGCCAGGCGGAACGGGCATCGGGTACTGGCG 36297

Qy        781 GTGGTGCGTGGCAGTGCGGTCAACCAGGACGGTGCGAGCAACGGTCTGACCGCGCCCAAC 840
Db      36298 GTTGTTCGTGGTAGCGCGGTGAATCAGGACGGTGCGTCGAACGGATTGACGGCCCCGAAC 36357

Qy        841 GGGCCGTCCCAGCAGCGTGTCATCCGCCAGGCCCTCGCCAACGCCGGCTTGTCGGCCGGT 900
Db      36358 GGGCTGGCCCAGGAGCGGGTCATTCAGCAGGTGCTCACGAGTGCGGGGCTGTCGGCGTCC 36417

Qy        901 GATGTCGACGCGGTGGAGGCCCACGGCACCGGCACCACTTTGGGCGACCCGATCGAGGCC 960
Db      36418 GATGTGGACGCTGTGGAGGCGCATGGAACGGGTACGCGGCTTGGTGATCCGATCGAGGCG 36477

Qy        961 CAGGCCCTCCTCGCGACCTACGGACAGGACCGTGCCGGCGAGGGGCCGCTGTGGCTGGGC 1020
Db      36478 CAGGCTCTGATAGCCGCCTATGGACAGGATCGGGACCGGGACCGGCCGCTGTGGTTGGGG 36537

Qy       1021 TCGGTCAAGTCCAATGTCGGTCACACACAGGCTGCCGCGGGCGTCGCCGGGGTGATCAAG 1080
Db      36538 TCGGTCAAGTCCAACATCGGTCATACGCAGGCGGCTGCGGGCGTCGCTGGTGTGATCAAG 36597

Qy       1081 ATGGTGATGGCGCTGCGGCATGGTCTGCTGCCGCGGACGTTGCATGTGGATGAGCCGTCG 1140
Db      36598 ATGGTCATGGCGATGCGGCACGGGGAGCTGCCGCGCACGTTGCACGTGGACGAGCCGAAT 36657

Qy       1141 CCGCATGTGGACTGGTCCGCGGGTGCGGTGCAGCTGCTGACGGAGACGGTGCCCTGGCCC 1200
Db      36658 TCGCACGTGGACTGGTCGGCTGGTGCGGTCCGACTCCTGACCGAGAACATCCGCTGGCCA 36717
```

```
Qy       1201 GGCGGGGAGGGGCGGCTACGGCGGGCAGGAGTGTCATCATTCGGCGTCAGCGGCACCAAC 1260
             ||       |  | |    || || || |||||||| || ||||| || |||||| ||||||
Db      36718 GG------GACGGGTACGCGCCGCGCTGGAGTGTCGTCGTTCGGGGTAAGCGGTACCAAC 36771

Qy       1261 GCCCACGTCATCCTCGAAGAAGCACC 1286
             || ||||||||||||||| | |  ||
Db      36772 GCACACGTCATCCTCGAACACGACCC 36797


RESULT 8
US-09-370-700-1
; Sequence 1, Application US/09370700
; Patent No. 6274350
; GENERAL INFORMATION:
;  APPLICANT: Baltz, Richard H
;  APPLICANT: Broughton, Mary C
;  APPLICANT: Crawford, Kathryn P
;  APPLICANT: Madduri, Krishnamurthy
;  APPLICANT: Treadway, Patti J
;  APPLICANT: Turner, Jan R
;  APPLICANT: Waldron, Clive
;  TITLE OF INVENTION: Biosynthetic Genes For Spinosyn Insecticide
;  FILE REFERENCE: 50489 DIV1
;  CURRENT APPLICATION NUMBER: US/09/370,700
;  CURRENT FILING DATE: 1999-08-09
;  EARLIER APPLICATION NUMBER: US 09/36987
;  EARLIER FILING DATE: 1998-03-09
;  NUMBER OF SEQ ID NOS: 39
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
;   LENGTH: 80161
;   TYPE: DNA
;   ORGANISM: Saccharopolyspora spinosa
US-09-370-700-1

  Query Match             55.4%;  Score 715.2;  DB 4;  Length 80161;
  Best Local Similarity   72.9%;  Pred. No. 1.8e-140;
  Matches  937;  Conservative    0;  Mismatches  343;  Indels    6;  Gaps    1;

Qy          1 GAGCCGATCGCCATCGTTGGCATGGCCTGTCGTTTCCCCGGCGGAGTGACCTCGGCGGAC 60
              || || ||||| ||||| |  |||| ||| || | ||| || || |||       ||||
Db      35518 GAACCCATCGCGATCGTGGCGATGGGCTGCCGGTACCCGGGTGGGGTGCAGGACCCGGAA 35577

Qy         61 GACTTCTGGGATCTGATCTCCTCCGAGCAGGACGCGATCGGCGGATTCCCCACCGACCGC 120
              |   | ||| | ||| || ||||||     ||||| ||||| | ||||||| | || ||
Db      35578 GGGCTGTGGAAACTGGTCGCCTCCGGTGGCGACGCCATCGGTGAATTCCCCGCTGATCGT 35637

Qy        121 GGCTGGGACCTGGACACGCTCTACGACCCCGACCCCGACCACCCCGGCACCTGCTACACC 180
              || ||| |||| |||  ||||||||| |||||||| || || ||||| ||||||||||||
Db      35638 GGTTGGCACCTCGACGAGCTCTACGATCCCGACCCGGATCAGCCCGGAACCTGCTACACC 35697

Qy        181 CGAAACGGCGGATTCCTCTACGACGCAGGCCACTTCGACGCCGAATTCTTCGGCATCAGC 240
              ||  ||||||| |||||| ||||||| ||| | |||||||| | ||||||| |||||||
Db      35698 CGGCACGGCGGCTTCCTCCACGACGCCGGCGAGTTCGACGCGGGATTCTTCGACATCAGC 35757

Qy        241 CCCCGCGAAGCCCTCGCCATGGACCCCCAGCAACGACTCCTCCTCGAAACCGCCTGGGAA 300
              ||||| || || ||||| |||||||| ||||| || || || || |||| | |||||||
Db      35758 CCCCGTGAGGCGCTCGCGATGGACCCGCAGCAGCGGCTGCTGCTGGAAATCTCCTGGGAG 35817

Qy        301 ACCATCGAACACGCCGGCATCAACCCCCACACCCTCCACGGCACCCCCACCGGAGTCTTC 360
              ||| |||||  ||| || ||  ||||     || | |  || | || |||||| || |||
Db      35818 ACCGTCGAATCCGCTGGGATGGACCCGAGGTCCTTGCGGGGGAGCCGCACCGGGGTGTTC 35877

Qy        361 ACCGGCACCAACGGACAGGACTACGCACTTCGCGTGCACAACGCGGGCCAGTCAACCGAT 420
              | ||     |      ||| ||| |      ||| |||   || ||  |       |||
Db      35878 GCGGGATTGATGTACGAGGGCTATGACACCGGCGCCCACCGGGCAGGAGAAGGTGTCGAA 35937

Qy        421 GGTTTCGCACTGACCGGAACCGCCGGCAGCGTCATCTCCGGTCGTATCTCGTACACGTTT 480
              || |       ||||| |  || || ||||||  ||| |||||  |  ||||  ||||
Db      35938 GGCTATCTCGGAACCGGCAATGCGGGAAGCGTCGCCTCTGGTCGGGTTGCGTATGCGTTC 35997

Qy        481 GGTTTTGAGGGTCCTGCGGTGTCGGTGGACACGGCTTGTTCCTCGTCGTTGGTGGCTTTG 540
              || || ||||| || |||||| |||| |||||||| || || ||||||||||||||  ||
```

```
Db      35998 GGGTTCGAGGGCCCAGCGGTGACGGTAGACACGGCGTGCTCGTCGTCGTTGGTGGCGCTG 36057

Qy        541 CATCTGGCCTGTCAGGCGTTGCGTGCGGGTGAGTGCTCGATGGCGCTTGCCGGGGGTGTG 600
              ||| |||| |||||| |||||||    ||| |||||      ||||||| ||||| || |||
Db      36058 CATTTGGCGTGTCAGTCGTTGCGGCAGGGCGAGTGTGATCTGGCGCTGGCCGGTGGAGTG 36117

Qy        601 ACGGTGATGTCGTCTCCGGGTGCCTTCGTGGAGTTTTCGCGGCAGCGGGGTCTGGCCGCG 660
              |||||||||||| | ||||     |||||||||| || || ||||| ||||  ||  ||
Db      36118 ACGGTGATGTCGACGCCGGAGAGGTTCGTGGAGTTCTCCCGTCAGCGTGGTCTCGCACCG 36177

Qy        661 GACGGGCATTGCAAGGCGTTCTCGGCGGCGGCGGACGGGACCGGCTGGGGTGAGGGTGTG 720
              || ||||  || ||| ||||| ||||||| ||||| || ||||| ||||||||||||||
Db      36178 GATGGGCGGTGTAAGTCGTTCGCGGCGGCTGCGGATGGAACCGGTTGGGGTGAGGGTGCC 36237

Qy        721 GGGATGCTGCTGGTGGAGCGGCTCTCCGACGCCCATCGCAACGGTCACCGTGTCCTGGCC 780
              ||  || || || |||||||||| || ||||||   || ||||| || || || |||||
Db      36238 GGTTTGGTGTTGCTGGAGCGGCTGTCAGACGCCAGGCGGAACGGGCATCGGGTACTGGCG 36297

Qy        781 GTGGTGCGTGGCAGTGCGGTCAACCAGGACGGTGCGAGCAACGGTCTGACCGCGCCCAAC 840
              || || ||||| || ||||| || ||||||||||||    |||||  |||| || || |||
Db      36298 GTTGTTCGTGGTAGCGCGGTGAATCAGGACGGTGCGTCGAACGGATTGACGGCCCCGAAC 36357

Qy        841 GGGCCGTCCCAGCAGCGTGTCATCCGCCAGGCCCTCGCCAACGCCGGCTTGTCGGCCGGT 900
              |||| | ||||| |||| ||||| |  ||||  ||| | |  || ||  |||||||
Db      36358 GGGCTGGCCCAGGAGCGGGTCATTCAGCAGGTGCTCACGAGTGCGGGGCTGTCGGCGTCC 36417

Qy        901 GATGTCGACGCGGTGGAGGCCCACGGCACCGGCACCACTTTGGGCGACCCGATCGAGGCC 960
              ||||| ||||| |||||||| || || || || ||      | || || ||||||||||
Db      36418 GATGTGGACGCTGTGGAGGCGCATGGAACGGGTACGCGGCTTGGTGATCCGATCGAGGCG 36477

Qy        961 CAGGCCCTCCTCGCGACCTACGGACAGGACCGTGCCGGCGAGGGGCCGCTGTGGCTGGGC 1020
              ||||| ||  | ||  |||| |||||||| || | | | ||  |||||||||| ||||
Db      36478 CAGGCTCTGATAGCCGCCTATGGACAGGATCGGGACCGGGACCGGCCGCTGTGGTTGGGG 36537

Qy       1021 TCGGTCAAGTCCAATGTCGGTCACACACAGGCTGCCGCGGGCGTCGCCGGGGTGATCAAG 1080
              |||||||||||||  ||||||| || ||||| || |||||||||| || |||||||||
Db      36538 TCGGTCAAGTCCAACATCGGTCATACGCAGGCGGCTGCGGGCGTCGCTGGTGTGATCAAG 36597

Qy       1081 ATGGTGATGGCGCTGCGGCATGGTCTGCTGCCGCGGACGTTGCATGTGGATGAGCCGTCG 1140
              ||||| |||||| ||||||| ||   ||||||||| |||||||| ||||| ||||||
Db      36598 ATGGTCATGGCGATGCGGCACGGGGAGCTGCCGCGCACGTTGCACGTGGACGAGCCGAAT 36657

Qy       1141 CCGCATGTGGACTGGTCCGCGGGTGCGGTGCAGCTGCTGACGGAGACGGTGCCCTGGCCC 1200
              ||||  |||||||||||| || |||||||| |  || ||||| ||||   | | ||||||
Db      36658 TCGCACGTGGACTGGTCGGCTGGTGCGGTCCGACTCCTGACCGAGAACATCCGCTGGCCA 36717

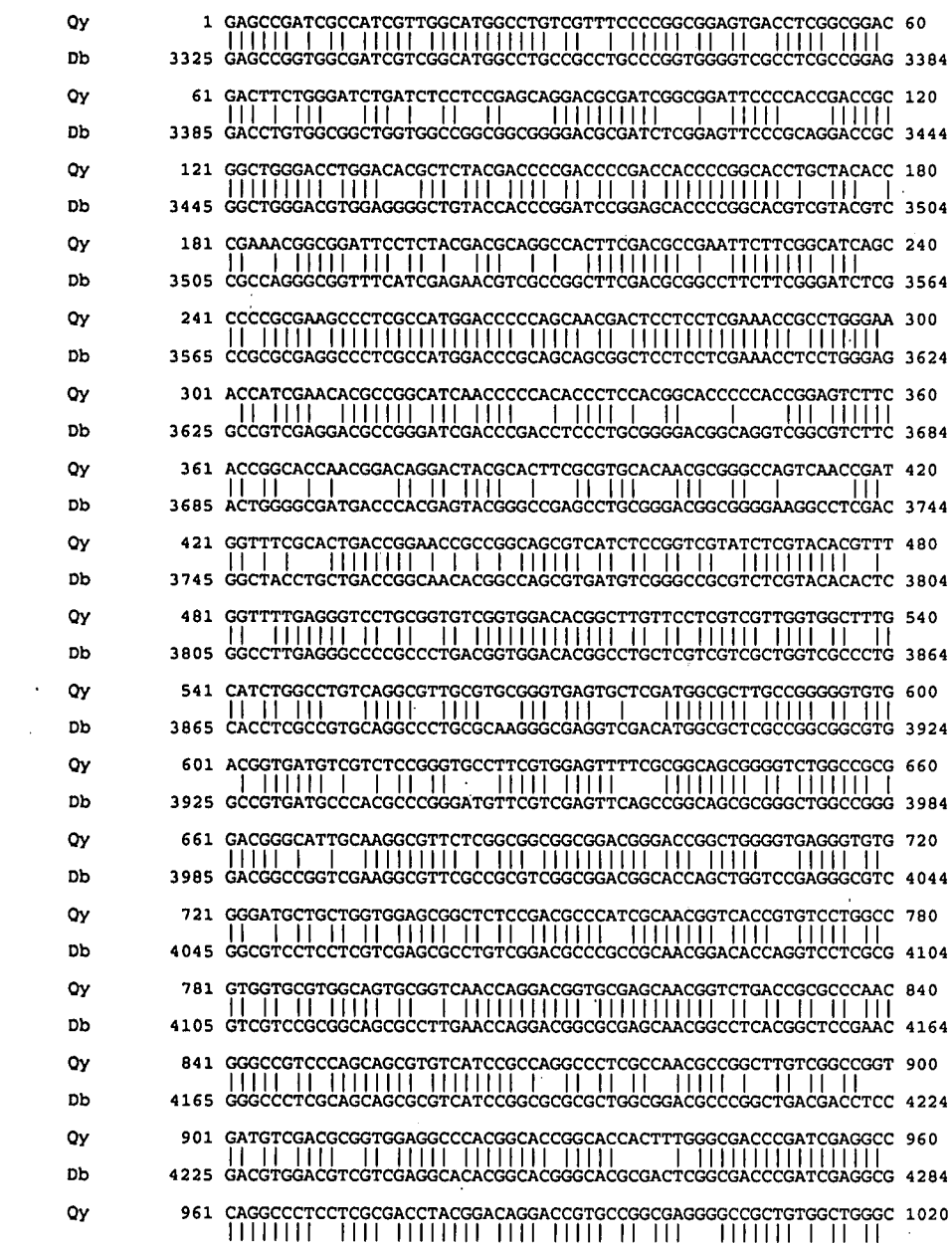
Qy       1201 GGCGGGGAGGGGCGGCTACGGCGGGCAGGAGTGTCATCATTCGGCGTCAGCGGCACCAAC 1260
              ||       |  | |    || || || |||||||| || ||||| || |||||| ||||||
Db      36718 GG------GACGGGTACGCGCCGCGCTGGAGTGTCGTCGTTCGGGGTAAGCGGTACCAAC 36771

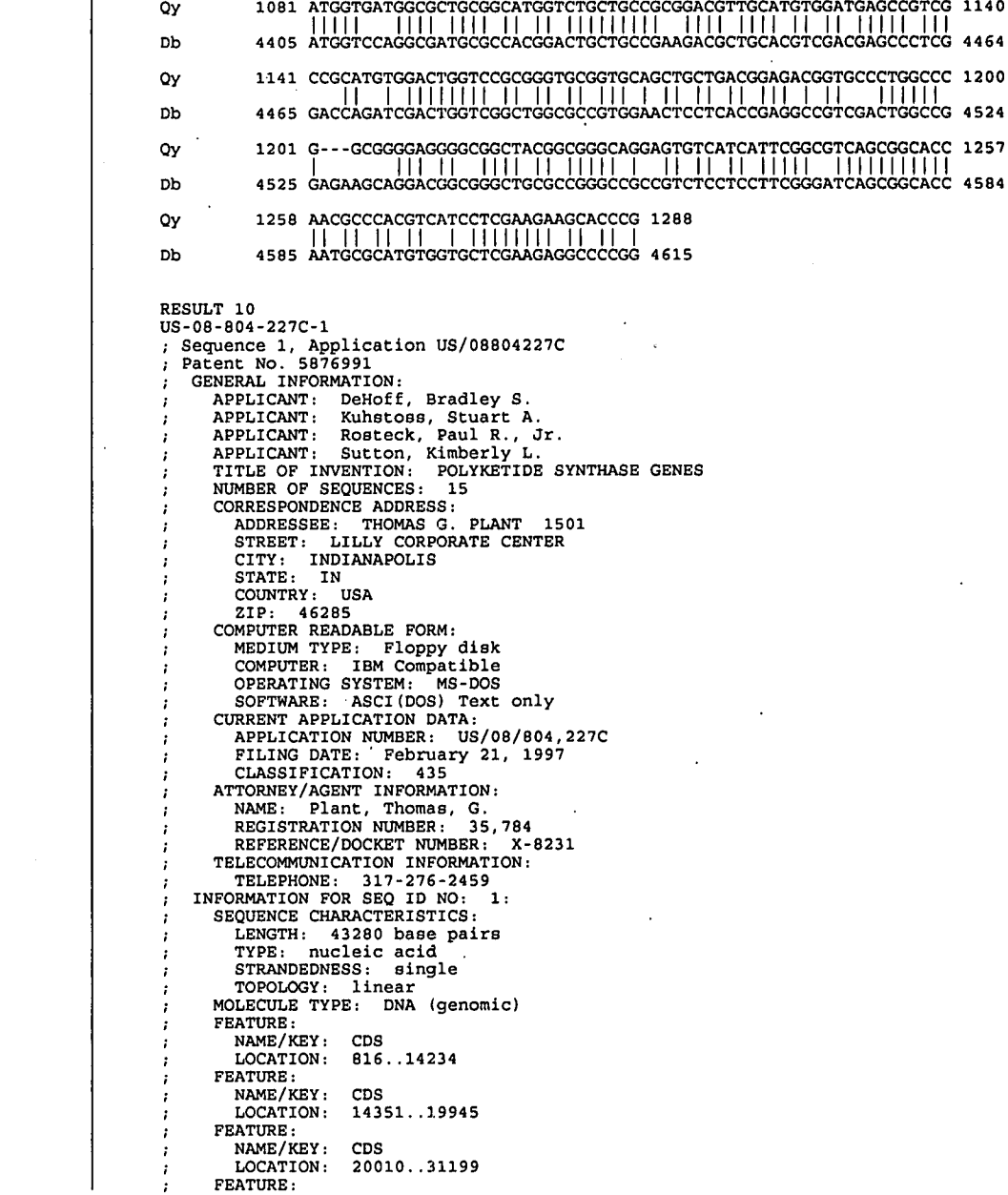
Qy       1261 GCCCACGTCATCCTCGAAGAAGCACC 1286
              || ||||||||||||||| | |  ||
Db      36772 GCACACGTCATCCTCGAACACGACCC 36797


RESULT 9
US-09-105-537-30
; Sequence 30, Application US/09105537A
; Patent No. 6265202
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT: Liu, H.
;  APPLICANT: Xue, Y.
;  APPLICANT: Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/105,537A
;  CURRENT FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 30
;   LENGTH: 13842
;   TYPE: DNA
```

; ORGANISM: Streptomyces venezuelae
US-09-105-537-30

Query Match 53.4%; Score 689.4; DB 4; Length 13842;
Best Local Similarity 71.4%; Pred. No. 3.2e-135;
Matches 922; Conservative 0; Mismatches 366; Indels 3; Gaps 1;





RESULT 10
US-08-804-227C-1
; Sequence 1, Application US/08804227C
; Patent No. 5876991
; GENERAL INFORMATION:
; APPLICANT: DeHoff, Bradley S.
; APPLICANT: Kuhstoss, Stuart A.
; APPLICANT: Rosteck, Paul R., Jr.
; APPLICANT: Sutton, Kimberly L.
; TITLE OF INVENTION: POLYKETIDE SYNTHASE GENES
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: THOMAS G. PLANT 1501
; STREET: LILLY CORPORATE CENTER
; CITY: INDIANAPOLIS
; STATE: IN
; COUNTRY: USA
; ZIP: 46285
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: ASCI(DOS) Text only
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/804,227C
; FILING DATE: February 21, 1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Plant, Thomas, G.
; REGISTRATION NUMBER: 35,784
; REFERENCE/DOCKET NUMBER: X-8231
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 317-276-2459
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 43280 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 816..14234
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 14351..19945
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 20010..31199
; FEATURE:

```
;      NAME/KEY:  CDS
;      LOCATION:  31232..36067
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  36249..41774
US-08-804-227C-1

  Query Match             53.4%;  Score 688.4;  DB 2;  Length 43280;
  Best Local Similarity   70.9%;  Pred. No. 6.3e-135;
  Matches  914;  Conservative    0;  Mismatches  376;  Indels    0;  Gaps    0;

Qy          1 GAGCCGATCGCCATCGTTGGCATGGCCTGTCGTTTCCCCGGCGGAGTGACCTCGGCGGAC 60
              |||||||||||| |||||   ||||||||| ||| | || || || || | | || | ||
Db      24675 GAGCCGATCGCGATCGTGTCCATGGCCTGCCGTATGCCGGGGGGTGTCAGCACGCCCGAG 24734

Qy         61 GACTTCTGGGATCTGATCTCCTCCGAGCAGGACGCGATCGGCGGATTCCCCACCGACCGC 120
              ||  | |||   ||| |       ||   ||||||||||||     |||||| || ||||
Db      24735 GAGCTGTGGCGGCTGCTGCGGGACGGCAAGGACGCGATCGGGCCGTTCCCCGCCAACCGG 24794

Qy        121 GGCTGGGACCTGGACACGCTCTACGACCCCGACCCCGACCACCCCGGCACCTGCTACACC 180
              |||||||||||||| | ||||||||||||||||| |||  |  ||||  |  |||
Db      24795 GGCTGGGACCTGGAGAACCTCTACGACCCCGACCCGGACGCCGACGGCCGCACCTATGTG 24854

Qy        181 CGAAACGGCGGATTCCTCTACGACGCAGGCCACTTCGACGCCGAATTCTTCGGCATCAGC 240
              ||  | ||||||||||||| |||| |||    |||||||| ||   ||||||||||||
Db      24855 CGCGAGGGCGGATTCCTCCACGAGGCACCGGACTTCGACCCCTCGTTCTTCGGCATCTCG 24914

Qy        241 CCCCGCGAAGCCCTCGCCATGGACCCCCAGCAACGACTCCTCCTCGAAACCGCCTGGGAA 300
              || ||||| || || || |||||||| ||||| || || || || || ||  |||||||
Db      24915 CCGCGCGAGGCGCTGGCGATGGACCCGCAGCAGCGGCTGCTGCTGGAGACTTCCTGGGAG 24974

Qy        301 ACCATCGAACACGCCGGCATCAACCCCCACACCCTCCACGGCACCCCCACCGGAGTCTTC 360
              || | || | ||||||||||·||||    |  || | ||||| ||  |||||  ||||||
Db      24975 GCCTTGGAGCGCGCCGGCATCGACCCGGCGAGGCTGCGCGGCAGCCGTACCGGCGTCTTC 25034

Qy        361 ACCGGCACCAACGGACAGGACTACGCACTTCGCGTGCACAACGCGGGCCAGTCAACCGAT 420
              |||| || ||||| ||| |||||    |  |    | || ||||  ||  |      |||
Db      25035 GTCGGGACGAACGGCCAGCACTACATGCCGCTGCTCCAGAACGGCGGGGACAGCTTCGAC 25094

Qy        421 GGTTTCGCACTGACCGGAACCGCCGGCAGCGTCATCTCCGGTCGTATCTCGTACACGTTT 480
              || |          ||||| | | |  | || || || || || || | ||||||  |||
Db      25095 GGCTATCTGGGCACCGGCAACTCGGCGAGTGTGATGTCGGGCCGGCTGTCGTACGTGTTC 25154

Qy        481 GGTTTTGAGGGTCCTGCGGTGTCGGTGGACACGGCTTGTTCCTCGTCGTTGGTGGCTTTG 540
              ||  | || || || || ||| | |||||||| || || ||  ||||  |||||||  ||
Db      25155 GGCCTCGAAGGCCCCGCCGTGACCGTGGACACCGCCTGCTCGGCGTCCCTGGTGGCGCTG 25214

Qy        541 CATCTGGCCTGTCAGGCGTTGCGTGCGGGTGAGTGCTCGATGGCGCTTGCCGGGGGTGTG 600
              || |||||·  ||||||  ||||    || ||||||   |||||||| | ||| || | |·
Db      25215 CACCTGGCGGTGCAGGCGATGCGGCGCGGCGAGTGCGACATGGCGCTGGTCGGCGGCGCG 25274

Qy        601 ACGGTGATGTCGTCTCCGGGTGCCTTCGTGGAGTTTTCGCGGCAGCGGGGTCTGGCCGCG 660
              ||||||||||| | || |       | |||||||| || ||||||||||    |  ||||
Db      25275 ACGGTGATGTCGACGCCCGAGATGCTGGTGGAGTTCTCCCGGCAGCGGGTGATCTCCGCC 25334

Qy        661 GACGGGCATTGCAAGGCGTTCTCGGCGGCGGCGGACGGGACCGGCTGGGGTGAGGGTGTG 720
              ||||  |  | ||| ||| | || |  || ||||     |      || ||||| |||
Db      25335 AACGGCCGGTCGAGGGCCTTCGCCGCCGGTGCCGACGGTGTGGCGCTCGGCGAGGGCGTG 25394

Qy        721 GGGATGCTGCTGGTGGAGCGGCTCTCCGACGCCCATCGCAACGGTCACCGTGTCCTGGCC 780
              ||  | |||||||||||||||| || |||||| | |||||||| || |  || |||||
Db      25395 GGCGTCCTGCTGGTGGAGCGGCTGTCGGACGCCGAGCGCAACGGCCATCCGGTGCTGGCG 25454

Qy        781 GTGGTGCGTGGCAGTGCGGTCAACCAGGACGGTGCGAGCAACGGTCTGACCGCGCCCAAC 840
              ||||| || |||   |||||||||||||||| ||    |||||| ||||| |||||||||
Db      25455 GTGGTCCGCGGCTCGGCGGTCAACCAGGACGGCGCCTCCAACGGGCTGACGGCGCCCAAC 25514

Qy        841 GGGCCGTCCCAGCAGCGTGTCATCCGCCAGGCCCTCGCCAACGCCGGCTTGTCGGCCGGT 900
              ||||| |||||||||||| || ||||| ||||| || ||  |||||||  ||  | |||
Db      25515 GGGCCCTCCCAGCAGCGGGTGATCCGGCAGGCGCTGGCGGACGCCGGGCTGCGGCCCGAG 25574

Qy        901 GATGTCGACGCGGTGGAGGCCCACGGCACCGGCACCACTTTGGGCGACCCGATCGAGGCC 960
              ||  ||||||| || ||||| ||||||||||||||    |||||||||| |||||||||

Db      25575 GACATCGACGCCGTCGAGGCGCACGGCACCGGCACCGAGCTGGGCGACCCCATCGAGGCC 25634

Qy        961 CAGGCCCTCCTCGCGACCTACGGACAGGACCGTGCCGGCGAGGGGCCGCTGTGGCTGGGC 1020
              |||| || |||||| ||||| |||  |  |||| | |  ||  | ||||||||||| |||
Db      25635 GAGGCGCTGCTCGCCACCTATGGAAGGACCCGTACGGCGGACCGCCCGCTGTGGCTCGGC 25694

Qy       1021 TCGGTCAAGTCCAATGTCGGTCACACACAGGCTGCCGCGGGCGTCGCCGGGGTGATCAAG 1080
              ||  | ||||||||  |||| ||||| ||||| ||||| ||||| || || || ||||||
Db      25695 TCCCTGAAGTCCAACATCGGGCACACCCAGGCCGCCGCCGGCGTGGCGGGCGTCATCAAG 25754

Qy       1081 ATGGTGATGGCGCTGCGGCATGGTCTGCTGCCGCGGACGTTGCATGTGGATGAGCCGTCG 1140
              |||||| | |||||| |  | |     |||||||  ||  |||| ||||||||||  |
Db      25755 ATGGTGCTCGCGCTGGGCAACGAGACACTGCCGCGCACCCTGCACGTGGATGAGCCCACA 25814

Qy       1141 CCGCATGTGGACTGGTCCGCGGGTGCGGTGCAGCTGCTGACGGAGACGGTGCCCTGGCCC 1200
              ||||  |||||||||||| | || |||||     ||||| || ||| |||||  ||||||
Db      25815 CCGCGCGTGGACTGGTCCTCTGGCGCGGTCTCCCTGCTCACCGAGCCGGTGGACTGGCCC 25874

Qy       1201 GGCGGGGAGGGGCGGCTACGGCGGGCAGGAGTGTCATCATTCGGCGTCAGCGGCACCAAC 1260
              | |||   |      ||  || || || |  ||||| || |||||| ||||||||||||||
Db      25875 GCCGGCCCGTCCGCGCCGCGCCGTGCGGCCGTGTCCTCGTTCGGCATCAGCGGCACCAAC 25934

Qy       1261 GCCCACGTCATCCTCGAAGAAGCACCCGCC 1290
              ||||||   ||||| ||  | || |||| |
Db      25935 GCCCACACGATCCTGGAGCAGGCCCCCGTC 25964


RESULT 11
US-07-642-734C-3
; Sequence 3, Application US/07642734C
; Patent No. 5824513
;  GENERAL INFORMATION:
;    APPLICANT:  Katz, L
;    APPLICANT:  Donadio, S
;    APPLICANT:  Mcalpine, J B
;    TITLE OF INVENTION:  Recombinant DNA Method for Producing
;    TITLE OF INVENTION:  Erythromycin Analogs
;    NUMBER OF SEQUENCES:  27
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Edward H. Gorman
;      STREET:  Abbott Laboratories D377/AP6D-2 One Abbott
;      STREET:  Park Rd
;      CITY:  Abbott Park
;      STATE:  IL
;      COUNTRY:  US
;      ZIP:  60064-3500
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/642,734C
;      FILING DATE:  17-JAN-91
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Danckers, Andreas M
;      REGISTRATION NUMBER:  32652
;      REFERENCE/DOCKET NUMBER:  4952.US.01
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  708-937-9396
;      TELEFAX:  708-938-2623
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  20235 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Saccharopolyspora erythraea
```

```
;       STRAIN:  NRRL 238
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  19..10722
;       OTHER INFORMATION:  /codon_start= 19
;       OTHER INFORMATION:  /function= "gene eryA"
;       OTHER INFORMATION:  /product= "eryA ORF2 encoding modules 3 & 4 for
;       OTHER INFORMATION:  6-deoxyerythronolide B"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  19..4470
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  module 3"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  97..1482
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  beta-ketoacyl ACP synthase of module 3"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  1693..2670
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  acyltransferase domain module 3"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  3406..3921
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  beta-ketoreductase domain of module 3"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  4171..4428
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  acyl carrier domain of module 3"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  4471..10722
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  module 4"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  4471..5847
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  beta-ketoacylACPsynhase domain of module"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  6054..7026
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  acyltransferase domain of module 4"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  7165..9216
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  dehydratase and enoylreductase domains m"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  9433..9984
;       OTHER INFORMATION:  /function= "approximate span
;       OTHER INFORMATION:  beta-ketoreductase of module 4"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  10225..10483
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  acyl carrier domain of module 4"
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  10723..20235
;       OTHER INFORMATION:  /codon_start= 10723
;       OTHER INFORMATION:  /function= "gene =eryA"
;       OTHER INFORMATION:  /product= "orf3 encoding modules 5 & 6
;       OTHER INFORMATION:  6-deoxyerythronolide B formatio"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  10723..15165
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  module 5"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  10831..12174
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  beta-ketoacylACPsynthase domain of modul"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  12379..13350
;       OTHER INFORMATION:  /function= "approximatr span of
;       OTHER INFORMATION:  acyltransferase domain of module 5"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  14062..14610
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  beta-ketoreductase of module 5"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  14857..15114
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  acyl carrier domain of module 5"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  15166..20235
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  module 6"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  15172..16569
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  beta-ketoacylACPsynthase domain of modul"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  16768..17721
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  acyltransferase domain of module 6"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  18379..18921
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  beta-ketoreductase domain of module 6"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  19149..19398
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  acyl carrier domain of module 6"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  19492..20235
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  thioesterase domain of module 6"
US-07-642-734C-3

  Query Match             53.1%;  Score 684.6;  DB 1;  Length 20235;
  Best Local Similarity   71.8%;  Pred. No. 3.4e-134;
  Matches  926;  Conservative    0;  Mismatches  354;  Indels    9;  Gaps    2;

Qy          1 GAGCCGATCGCCATCGTTGGCATGGCCTGTCGTTTCCCCGGCGGAGTGACCTCGGCGGAC 60
              ||||| |||||||||| ||||| | ||| |||||||||||||   |   |||| | ||
Db       4483 GAGCCCATCGCCATCGTCGGCATCGGCTGCCGTTTCCCCGGCGGCATCGGCTCGCCCGAG 4542

Qy         61 GACTTCTGGGATCTGATCTCCTCCGAGCAGGACGCGATCGGCGGATTCCCCACCGACCGC 120
               | || |||    || |  ||   |   | ||   | |  ||| |||||  |||||||
Db       4543 CAGTTGTGGCGGGTGCTGGCCGAGGGCGCGAACCTCACCACCGGCTTCCCGGCCGACCGG 4602

Qy        121 GGCTGGGACCTGGACACGCTCTACGACCCCGACCCCGACCACCCCGGCACCTGCTACACC 180
              ||||||||| | |    |||||||| |||| ||||| ||| ||||||||||| |||||
Db       4603 GGCTGGGACATCGGGCGGCTCTACCACCCGGACCCGGACAACCCCGGCACCAGCTACGTG 4662

Qy        181 CGAAACGGCGGATTCCTCTACGACGCAGGCCACTTCGACGCCGAATTCTTCGGCATCAGC 240
              || ||||| ||||||  |||||| |   | |||||| | |  ||||||||||||
Db       4663 GACAAGGGCGGGTTCCTCACCGACGCGGCGGATTTCGACCCGGGCTTCTTCGGCATCACG 4722
```

```
Qy       241 CCCCGCGAAGCCCTCGCCATGGACCCCCAGCAACGACTCCTCCTCGAAACCGCCTGGGAA 300
            |||||||||||  ||  ||  ||||||||  |||||  ||  |||  |  ||  ||  ||  ||  |||||
Db      4723 CCCCGCGAAGCGCTGGCGATGGACCCGCAGCAGCGCCTCATGCTGGAGACGGCGTGGGAG 4782

Qy       301 ACCATCGAACACGCCGGCATCAACCCCCACACCCTCCACGGCACCCCCACCGGAGTCTTC 360
              |  |  ||||  |||  ||||||  |||||  ||  ||||  |  ||||||  ||||||  ||||||
Db      4783 GCAGTGGAACGCGCGGGCATCGACCCCGACGCCCTGCGAGGCACCGACACCGGCGTCTTC 4842

Qy       361 ACCGGCACCAACGGACAGGACTACGCACTTCGCGTGCACAACGCGGGCCAGTCAACCGAT 420
               |||||  |||||  |||  ||||    |  |    ||  |    |  ||  |  |       |||
Db      4843 GTCGGCATGAACGGCCAGTCCTACATGCAGCTGCTGGCCGGTGAGGCCGAACGCGTCGAC 4902

Qy       421 GGTTTCGCACTGACCGGAACCGCCGGCAGCGTCATCTCCGGTCGTATCTCGTACACGTTT 480
            ||  |  |        |||||  |  |||   |||||   |||||||  ||  |||  |  |||||  ||
Db      4903 GGCTACCAGGGCCTCGGAAACTCCGCGAGCGTGCTCTCCGGGCGCATCGCCTACACCTTC 4962

Qy       481 GGTTTTGAGGGTCCTGCGGTGTCGGTGGACACGGCTTGTTCCTCGTCGTTGGTGGCTTTG 540
            ||  |  |||||  ||  |||  ||  ||||||||||  ||  ||  ||  ||  |||  ||||  |    |
Db      4963 GGCTGGGAGGGCCCGGCGCTGACGGTGGACACCGCGTGCTCGTCCTCGCTGGTCGGCATC 5022

Qy       541 CATCTGGCCTGTCAGGCGTTGCGTGCGGGTGAGTGCTCGATGGCGCTTGCCGGGGGTGTG 600
            ||  ||  ||     ||||||  ||||     ||||||||||  |||||||  |||||  ||  ||
Db      5023 CACCTCGCGATGCAGGCGCTGCGGCGCGGTGAGTGCTCCCTGGCGCTGGCCGGCGGCGTC 5082

Qy       601 ACGGTGATGTCGTCTCCGGGTGCCTTCGTGGAGTTTTCGCGGCAGCGGGGTCTGGCCGCG 660
            |||||  |||||.    |||     |||||||  ||  ||       ||||||  ||  ||  |||  |
Db      5083 ACGGTCATGTCCGACCCGTACACCTTCGTCGACTTCAGCACGCAGCGCGGGCTCGCCTCC 5142

Qy       661 GACGGGCATTGCAAGGCGTTCTCGGCGGCGGCGGACGGGACCGGCTGGGGTGAGGGTGTG 720
            |||||  |  ||||||||||||||  |||   |||  |||||     ||      |     ||  ||  ||
Db      5143 GACGGTCGCTGCAAGGCGTTCTCCGCGCGGGCCGACGGCTTCGCGCTGTCGGAAGGCGTC 5202

Qy       721 GGGATGCTGCTGGTGGAGCGGCTCTCCGACGCCCATCGCAACGGTCACCGTGTCCTGGCC 780
            |     ||||  ||  ||||||  |||  |||    ||  |     ||||||  ||||    ||  ||||||
Db      5203 GCCGCGCTGGTGCTGGAGCCGCTTTCCCGGGCGCGCGCCAACGGGCACCAGGTGCTGGCC 5262

Qy       781 GTGGTGCGTGGCAGTGCGGTCAACCAGGACGGTGCGAGCAACGGTCTGACCGCGCCCAAC 840
            |||  ||||  |||||  ||||||||||||||||||||  ||||||||||   ||||  ||||||
Db      5263 GTGCTGCGCGGCAGCGCGGTCAACCAGGACGGTGCCAGCAACGGTCTCGCCGCTCCCAAC 5322

Qy       841 GGGCCGTCCCAGCAGCGTGTCATCCGCCAGGCCCTCGCCAACGCCGGCTTGTCGGCCGGT 900
            ||  |||||  |||  ||||  ||  |||||  |||||  ||||||      |  |||  ||  ||||||
Db      5323 GGCCCGTCGCAGGAGCGGGTGATCCGGCAGGCGCTCGCCGCTTCGGGCGTGCCGGCCGCG 5382

Qy       901 GATGTCGACGCGGTGGAGGCCCACGGCACCGGCACCACTTTGGGCGACCCGATCGAGGCC 960
            ||  |||||||  |||||||  |||||  ||  ||||||      |  ||||||||||||||||||
Db      5383 GACGTCGACGTCGTGGAGGCGCACGGGACGGGCACCGAGCTCGGCGACCCGATCGAGGCC 5442

Qy       961 CAGGCCCTCCTCGCGACCTACGGACAGGACCGTGCCGGCGAGGGGCCGCTGTGGCTGGGC 1020
               ||  |||  ||||||||||||  |||||||       ||||  ||||||||  ||||  |||
Db      5443 GGCGCGCTCATCGCGACCTACGGCCAGGACC------GCGACCGGCCGCTGCGGCTCGGC 5496

Qy       1021 TCGGTCAAGTCCAATGTCGGTCACACACAGGCTGCCGCGGGCGTCGCCGGGGTGATCAAG 1080
              |||||  |||  ||||  ||||  |||||  |||||  ||  |||||||  |||  ||  |||||||||
Db      5497 TCGGTGAAGACCAACATCGGCCACACCCAGGCCGCGGCGGGCGCCGCGGGCGTGATCAAG 5556

Qy       1081 ATGGTGATGGCGCTGCGGCATGGTCTGCTGCCGCGGACGTTGCATGTGGATGAGCCGTCG 1140
               |  |||  |||||  |||||||  ||   |||||||  |||  |||||||  |   ||  ||||  |||
Db      5557 GTCGTGCTGGCGATGCGGCACGGGATGCTGCCCCGGTCGTTGCACGCCGACGAGCTGTCC 5616

Qy       1141 CCGCATGTGGACTGGTCCGCGGGTGCGGTGCAGCTGCTGACGGAGACGGTGCCCTGGCCC 1200
              |||||  |  ||||||    ||||  ||  |||  ||  |||||    |||  ||||||  |||||
Db      5617 CCGCACATCGACTGGGAGTCGGGGGCCGTGGAGGTGCTGCGCGAGGAGGTGCCGTGGCCG 5676

Qy       1201 GGCGGGGAGGGGCGGCTACGGCGGGCAGGAGTGTCATCATTCGGCGTCAGCGGCACCAAC 1260
              |  ||  ||    |||  |  ||||||||  ||  |||||  ||  ||||||||||||||  ||||||
Db      5677 GCGGGTGA---GCGCCCCCGGCGGGCGGGGGTGTCGTCCTTCGGCGTCAGCGGAACCAAC 5733

Qy       1261 GCCCACGTCATCCTCGAAGAAGCACCCGC 1289
              ||  |||||  |||  |||||||  |||||  ||
Db      5734 GCGCACGTGATCGTCGAAGAGGCACCAGC 5762
```

```
RESULT 12
US-08-439-009A-3
; Sequence 3, Application US/08439009A
; Patent No. 6004787
;  GENERAL INFORMATION:
;    APPLICANT:  Donadio, S
;    APPLICANT:  Katz, L
;    APPLICANT:  Mcalpine, J B
;    TITLE OF INVENTION:  Method of Directing Biosynthesis of
;    TITLE OF INVENTION:  Specific Polyketides
;    NUMBER OF SEQUENCES:  27
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Steven F. Weinstock
;      STREET:  Abbott Laboratories D377/AP6D-2 One Abbott
;      STREET:  Park Rd
;      CITY:  Abbott Park
;      STATE:  IL
;      COUNTRY:  US
;      ZIP:  60064-3500
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/439,009A
;      FILING DATE:  11-MAY-1995
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Casuto, Dianne
;      REGISTRATION NUMBER:  40,943
;      REFERENCE/DOCKET NUMBER:  4952.US.D1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  847-938-3137
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  20235 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Saccharopolyspora erythraea
;      STRAIN:  NRRL 238
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  19..10722
;      OTHER INFORMATION:  /codon_start= 19
;      OTHER INFORMATION:  /function= "gene eryA"
;      OTHER INFORMATION:  /product= "eryA ORF2 encoding modules 3 & 4 for
;      OTHER INFORMATION:  6-deoxyerythronolide B"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  19..4470
;      OTHER INFORMATION:  /function= "approximate span of
;      OTHER INFORMATION:  module 3"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  97..1482
;      OTHER INFORMATION:  /function= "approximate span of
;      OTHER INFORMATION:  beta-ketoacyl ACP synthase of module 3"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  1693..2670
;      OTHER INFORMATION:  /function= "approximate span of
;      OTHER INFORMATION:  acyltransferase domain module 3"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  3406..3921
;      OTHER INFORMATION:  /function= "approximate span of
```

```
;       OTHER INFORMATION:  beta-ketoreductase domain of module 3"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  4171..4428
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  acyl carrier domain of module 3"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  4471..10722
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  module 4"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  4471..5847
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  beta-ketoacylACPsynhase domain of module"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  6054..7026
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  acyltransferase domain of module 4"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  7165..9216
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  dehydratase and enoylreductase domains m"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  9433..9984
;       OTHER INFORMATION:  /function= "approximate span
;       OTHER INFORMATION:  beta-ketoreductase of module 4"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  10225..10483
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  acyl carrier domain of module 4"
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  10723..20235
;       OTHER INFORMATION:  /codon_start= 10723
;       OTHER INFORMATION:  /function= "gene =eryA"
;       OTHER INFORMATION:  /product= "orf3 encoding modules 5 & 6
;       OTHER INFORMATION:  6-deoxyerythronolide B formatio"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  10723..15165
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  module 5"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  10831..12174
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  beta-ketoacylACPsynthase domain of modul"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  12379..13350
;       OTHER INFORMATION:  /function= "approximatr span of
;       OTHER INFORMATION:  acyltransferase domain of module 5"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  14062..14610
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  beta-ketoreductase of module 5"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  14857..15114
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  acyl carrier domain of module 5"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  15166..20235
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  module 6"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  15172..16569
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  beta-ketoacylACPsynthase domain of modul"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  16768..17721
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  acyltransferase domain of module 6"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  18379..18921
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  beta-ketoreductase domain of module 6"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  19149..19398
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  acyl carrier domain of module 6"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  19492..20235
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  thioesterase domain of module 6"
US-08-439-009A-3

  Query Match             53.1%;  Score 684.6;  DB 3;  Length 20235;
  Best Local Similarity   71.8%;  Pred. No. 3.4e-134;
  Matches  926;  Conservative    0;  Mismatches  354;  Indels    9;  Gaps    2;

Qy          1 GAGCCGATCGCCATCGTTGGCATGGCCTGTCGTTTCCCCGGCGGAGTGACCTCGGCGGAC. 60
              ||||| |||||||||||| ||||| | ||| ||||||||||||||  |   |||| | ||
Db       4483 GAGCCCATCGCCATCGTCGGCATCGGCTGCCGTTTCCCCGGCGGCATCGGCTCGCCCGAG 4542

Qy         61 GACTTCTGGGATCTGATCTCCTCCGAGCAGGACGCGATCGGCGGATTCCCCACCGACCGC 120
               || |||    || | ||    |     | ||   | |  ||| |||||  |||||||
Db       4543 CAGTTGTGGCGGGTGCTGGCCGAGGGCGCGAACCTCACCACCGGCTTCCCGGCCGACCGG 4602

Qy        121 GGCTGGGACCTGGACACGCTCTACGACCCCGACCCCGACCACCCCGGCACCTGCTACACC 180
              |||||||||  | |    |||||||  |||| |||||  ||| ||||||||||| |||||
Db       4603 GGCTGGGACATCGGGCGGCTCTACCACCCGGACCCGGACAACCCCGGCACCAGCTACGTG 4662

Qy        181 CGAAACGGCGGATTCCTCTACGACGCAGGCCACTTCGACGCCGAATTCTTCGGCATCAGC 240
              ||  |||||  ||||||  ||||||  |   | ||||||  | |  ||||||||||||
Db       4663 GACAAGGGCGGGTTCCTCACCGACGCGGCGGATTTCGACCCGGGCTTCTTCGGCATCACG 4722

Qy        241 CCCCGCGAAGCCCTCGCCATGGACCCCCAGCAACGACTCCTCCTCGAAACCGCCTGGGAA 300
              |||||||||| || || |||||||| ||||| || ||| | || || || || ||||| |
Db       4723 CCCCGCGAAGCGCTGGCGATGGACCCGCAGCAGCGCCTCATGCTGGAGACGGCGTGGGAG 4782

Qy        301 ACCATCGAACACGCCGGCATCAACCCCCACACCCTCCACGGCACCCCCCACCGGAGTCTTC 360
              |  | |||| ||| |||||| ||||| || |||| |  ||||||  ||||||  ||||||
Db       4783 GCAGTGGAACGCGCGGGCATCGACCCCGACGCCCTGCGAGGCACCGACACCGGCGTCTTC 4842

Qy        361 ACCGGCACCAACGGACAGGACTACGCACTTCGCGTGCACAACGCGGGCCAGTCAACCGAT 420
              |||||  ||||| ||| ||||   |  |   ||  |   | || | |       |||
Db       4843 GTCGGCATGAACGGCCAGTCCTACATGCAGCTGCTGGCCGGTGAGGCCGAACGCGTCGAC 4902

Qy        421 GGTTTCGCACTGACCGGAACCGCCGGCAGCGTCATCTCCGGTCGTATCTCGTACACGTTT 480
              || | |      ||||| | |||  |||||  ||||||| || ||| | ||||| ||
Db       4903 GGCTACCAGGGCCTCGGAAACTCCGCGAGCGTGCTCTCCGGGCGCATCGCCTACACCTTC 4962

Qy        481 GGTTTTGAGGGTCCTGCGGTGTCGGTGGACACGGCTTGTTCCTCGTCGTTGGTGGCTTTG 540
              || |  ||||| || ||| || |||||||||| || || || || ||| |||| |   |
Db       4963 GGCTGGGAGGGCCCGGCGCTGACGGTGGACACCGCGTGCTCGTCCTCGCTGGTCGGCATC 5022

Qy        541 CATCTGGCCTGTCAGGCGTTGCGTGCGGGTGAGTGCTCGATGGCGCTTGCCGGGGGTGTG 600
              || || ||    |||||| ||||    |||||||||||  |||||| ||||| || ||
Db       5023 CACCTCGCGATGCAGGCGCTGCGGCGCGGTGAGTGCTCCCTGGCGCTGGCCGGCGGCGTC 5082

Qy        601 ACGGTGATGTCGTCTCCGGGTGCCTTCGTGGAGTTTTCGCGGCAGCGGGGTCTGGCCGCG 660
              ||||| |||||    |||    ||||||| || ||      |||||| || || ||| |
```

```
Db       5083 ACGGTCATGTCCGACCCGTACACCTTCGTCGACTTCAGCACGCAGCGCGGGCTCGCCTCC 5142

Qy        661 GACGGGCATTGCAAGGCGTTCTCGGCGGCGGCGGACGGGACCGGCTGGGGTGAGGGTGTG 720

Db       5143 GACGGTCGCTGCAAGGCGTTCTCCGCGCGGGCCGACGGCTTCGCGCTGTCGGAAGGCGTC 5202

Qy        721 GGGATGCTGCTGGTGGAGCGGCTCTCCGACGCCCATCGCAACGGTCACCGTGTCCTGGCC 780

Db       5203 GCCGCGCTGGTGCTGGAGCCGCTTTCCCGGGCGCGCGCCAACGGGCACCAGGTGCTGGCC 5262

Qy        781 GTGGTGCGTGGCAGTGCGGTCAACCAGGACGGTGCGAGCAACGGTCTGACCGCGCCCAAC 840

Db       5263 GTGCTGCGCGGCAGCGCGGTCAACCAGGACGGTGCCAGCAACGGTCTCGCCGCTCCCAAC 5322

Qy        841 GGGCCGTCCCAGCAGCGTGTCATCCGCCAGGCCCTCGCCAACGCCGGCTTGTCGGCCGGT 900

Db       5323 GGCCCGTCGCAGGAGCGGGTGATCCGGCAGGCGCTCGCCGCTTCGGGCGTGCCGGCCGCG 5382

Qy        901 GATGTCGACGCGGTGGAGGCCCACGGCACCGGCACCACTTTGGGCGACCCGATCGAGGCC 960

Db       5383 GACGTCGACGTCGTGGAGGCGCACGGGACGGGCACCGAGCTCGGCGACCCGATCGAGGCC 5442

Qy        961 CAGGCCCTCCTCGCGACCTACGGACAGGACCGTGCCGGCGAGGGGCCGCTGTGGCTGGGC 1020

Db       5443 GGCGCGCTCATCGCGACCTACGGCCAGGACC------GCGACCGGCCGCTGCGGCTCGGC 5496

Qy       1021 TCGGTCAAGTCCAATGTCGGTCACACACAGGCTGCCGCGGGCGTCGCCGGGGTGATCAAG 1080

Db       5497 TCGGTGAAGACCAACATCGGCCACACCCAGGCCGCGGCGGGCGCCGCGGGCGTGATCAAG 5556

Qy       1081 ATGGTGATGGCGCTGCGGCATGGTCTGCTGCCGCGGACGTTGCATGTGGATGAGCCGTCG 1140

Db       5557 GTCGTGCTGGCGATGCGGCACGGGATGCTGCCCCGGTCGTTGCACGCCGACGAGCTGTCC 5616

Qy       1141 CCGCATGTGGACTGGTCCGCGGGTGCGGTGCAGCTGCTGACGGAGACGGTGCCCTGGCCC 1200

Db       5617 CCGCACATCGACTGGGAGTCGGGGGCCGTGGAGGTGCTGCGCGAGGAGGTGCCGTGGCCG 5676

Qy       1201 GGCGGGGAGGGGCGGCTACGGCGGGCAGGAGTGTCATCATTCGGCGTCAGCGGCACCAAC 1260

Db       5677 GCGGGTGA---GCGCCCCCGGCGGGCGGGGGTGTCGTCCTTCGGCGTCAGCGGAACCAAC 5733

Qy       1261 GCCCACGTCATCCTCGAAGAAGCACCCGC 1289

Db       5734 GCGCACGTGATCGTCGAAGAGGCACCAGC 5762


RESULT 13
US-09-105-537-1
; Sequence 1, Application US/09105537A
; Patent No. 6265202
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT: Liu, H.
;  APPLICANT: Xue, Y.
;  APPLICANT: Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/105,537A
;  CURRENT FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 1
;   LENGTH: 15872
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-105-537-1

  Query Match             52.7%;  Score 680.4;  DB 4;  Length 15872;
  Best Local Similarity   71.0%;  Pred. No. 2.5e-133;
  Matches  916;  Conservative    0;  Mismatches  371;  Indels    3;  Gaps    1;

Qy          1 GAGCCGATCGCCATCGTTGGCATGGCCTGTCGTTTCCCCGGCGGAGTGACCTCGGCGGAC 60

Db       8498 GATCCGATCGCGATCGTGGGCATGGCATGCCGCTATCCGGGTGGTGTGTCGTCGCCGGAG 8557

Qy         61 GACTTCTGGGATCTGATCTCCTCCGAGCAGGACGCGATCGGCGGATTCCCCACCGACCGC 120

Db       8558 GACCTGTGGCGGCTGGTGGCCGAGGGGACGGACGCGATCAGCGAGTTCCCCGTCAACCGC 8617

Qy        121 GGCTGGGACCTGGACACGCTCTACGACCCCGACCCCGACCACCCCGGCACCTGCTACACC 180

Db       8618 GGCTGGGACCTGGAGAGCCTCTACGACCCGGATCCCGAGTCGAAGGGCACCACGTACTGC 8677

Qy        181 CGAAACGGCGGATTCCTCTACGACGCAGGCCACTTCGACGCCGAATTCTTCGGCATCAGC 240

Db       8678 CGGGAGGGCGGGTTCCTGGAAGGCGCCGGTGACTTCGACGCCGCCTTCTTCGGCATCTCG 8737

Qy        241 CCCCGCGAAGCCCTCGCCATGGACCCCCAGCAACGACTCCTCCTCGAAACCGCCTGGGAA 300

Db       8738 CCGCGCGAGGCCCTGGTGATGGACCCGCAGCAGCGGCTGCTGCTGGAGGTGTCCTGGGAG 8797

Qy        301 ACCATCGAACACGCCGGCATCAACCCCCACACCCTCCACGGCACCCCCACCGGAGTCTTC 360

Db       8798 GCGCTGGAACGCGCGGGCATCGACCCGTCCTCGCTGCGCGGCAGCCGCGGTGGTGTCTAC 8857

Qy        361 ACCGGCACCAACGGACAGGACTACGCACTTCGCGTGCACAACGCGGGCCAGTCAACCGAT 420

Db       8858 GTGGGCGCCGCGCACGGCTCGTACGCCTCCGATCCCCGGCTGGTGCCCGAGGGCTCGGAG 8917

Qy        421 GGTTTCGCACTGACCGGAACCGCCGGCAGCGTCATCTCCGGTCGTATCTCGTACACGTTT 480

Db       8918 GGCTATCTGCTGACCGGCAGCGCCGACGCGGTGATGTCCGGCCGCATCTCCTACGCGCTC 8977

Qy        481 GGTTTTGAGGGTCCTGCGGTGTCGGTGGACACGGCTTGTTCCTCGTCGTTGGTGGCTTTG 540

Db       8978 GGTCTCGAAGGACCGTCCATGACGGTGGAGACGGCCTGCTCCTCCTCGCTGGTGGCGCTG 9037

Qy        541 CATCTGGCCTGTCAGGCGTTGCGTGCGGGTGAGTGCTCGATGGCGCTTGCCGGGGGTGTG 600

Db       9038 CATCTGGCGGTACGGGCGCTGCGGCACGGCGAGTGCGGGCTCGCGCTGGCGGGCGGGGTG 9097

Qy        601 ACGGTGATGTCGTCTCCGGGTGCCTTCGTGGAGTTTTCGCGGCAGCGGGGTCTGGCCGCG 660

Db       9098 GCGGTGATGGCCGATCCGGCGGCGTTCGTGGAGTTCTCCCGGCAGAAGGGGCTGGCCGCC 9157

Qy        661 GACGGGCATTGCAAGGCGTTCTCGGCGGCGGCGGACGGGACCGGCTGGGGTGAGGGTGTG 720

Db       9158 GACGGCCGCTGCAAGGCGTTCTCGGCCGCCGCCGACGGCACCGGCTGGGCCGAGGGCGTC 9217

Qy        721 GGGATGCTGCTGGTGGAGCGGCTCTCCGACGCCCATCGCAACGGTCACCGTGTCCTGGCC 780

Db       9218 GGCGTGCTCGTCCTGGAGCGGCTGTCGGACGCGCGCCGCGCGGGGCACACGGTCCTCGGC 9277

Qy        781 GTGGTGCGTGGCAGTGCGGTCAACCAGGACGGTGCGAGCAACGGTCTGACCGCGCCCAAC 840

Db       9278 CTGGTCACCGGCACCGCGGTCAACCAGGACGGTGCCTCCAACGGGCTGACCGCGCCCAAC 9337

Qy        841 GGGCCGTCCCAGCAGCGTGTCATCCGCCAGGCCCTCGCCAACGCCGGCTTGTCGGCCGGT 900

Db       9338 GGCCCAGCCCAGCAACGCGTCATCGCCGAGGCGCTCGCCGACGCCGGGCTGTCCCCGGAG 9397

Qy        901 GATGTCGACGCGGTGGAGGCCCACGGCACCGGCACCACTTTGGGCGACCCGATCGAGGCC 960

Db       9398 GACGTGGACGCGGTCGAGGCGCACGGCACCGGCACCCGGCTCGGCGACCCCATCGAGGCC 9457

Qy        961 CAGGCCCTCCTCGCGACCTACGGACAGGACCGTGCCGGCGAGGGGCCGCTGTGGCTGGGC 1020

Db       9458 GGGGCGCTGCTCGCCGCCTCCGGACGGAACCGTTCCGGCGACCACCCGCTGTGGCTCGGC 9517

Qy       1021 TCGGTCAAGTCCAATGTCGGTCACACACAGGCTGCCGCGGGCGTCGCCGGGGTGATCAAG 1080

Db       9518 TCGCTGAAGTCCAACATCGGGCATGCCCAGGCCGCCGCCGGTGTCGGCGGCGTCATCAAG 9577

Qy       1081 ATGGTGATGGCGCTGCGGCATGGTCTGCTGCCGCGGACGTTGCATGTGGATGAGCCGTCG 1140
```

```
Db       9578 ATGCTCCAGGCGCTGCGGCACGGCTTGCTGCCCCGCACCCTCCACGCCGACGAGCCGACC 9637

Qy       1141 CCGCATGTGGACTGGTCCGCGGGTGCGGTGCAGCTGCTGACGGAGACGGTGCCCTGGCCC 1200
              ||||||| ||||||  | | ||   ||| | |||||| ||       |||||| ||||
Db       9638 CCGCATGCCGACTGGAGCTCCGGCCGGGTACGGCTGCTCACCTCCGAGGTGCCGTGGC-- 9695

Qy       1201 GGCGGGGAGGGGCGGCTACGGCGGGCAGGAGTGTCATCATTCGGCGTCAGCGGCACCAAC 1260
                ||   || ||||  |||||| | || |||||  | |||||||| ||||||||||
Db       9696 -AGCGGACCGGCCGGCCCCGGCGGACCGGGGTGTCCGCCTTCGGCGTCGGCGGCACCAAT 9754

Qy       1261 GCCCACGTCATCCTCGAAGAAGCACCCGCC 1290
              ||||| ||| ||||||||||  |||||||||
Db       9755 GCCCATGTCGTCCTCGAAGAGGCACCCGCC 9784


RESULT 14
US-09-105-537-36
; Sequence 36, Application US/09105537A
; Patent No. 6265202
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT: Liu, H.
;  APPLICANT: Xue, Y.
;  APPLICANT: Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/105,537A
;  CURRENT FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 36
;   LENGTH: 4041
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-105-537-36

  Query Match             52.2%;  Score 673;  DB 4;  Length 4041;
  Best Local Similarity   70.1%;  Pred. No. 6.7e-132;
  Matches  904;  Conservative    0;  Mismatches  385;  Indels    0;  Gaps    0;

Qy          1 GAGCCGATCGCCATCGTTGGCATGGCCTGTCGTTTCCCCGGCGGAGTGACCTCGGCGGAC 60
              ||||| || || ||||| ||||||  ||| || ||| | |||||| |     ||  | ||
Db        103 GAGCCCATGGCGATCGTCGGCATGAGCTGCCGGTTCGCGGGCGGAATCCGGTCCCCCGAG 162

Qy         61 GACTTCTGGGATCTGATCTCCTCCGAGCAGGACGCGATCGGCGGATTCCCCACCGACCGC 120
              ||| |||||||     || || | |   ||||  | ||  ||   | ||    || |||
Db        163 GACCTCTGGGACGCCGTCGCCGCGGGCAAGGACCTGGTCTCCGAGGTACCGGAGGAGCGC 222

Qy        121 GGCTGGGACCTGGACACGCTCTACGACCCCGACCCCGACCACCCCGGCACCTGCTACACC 180
              |||||||| | ||| | |||||||||| |  ||||  | |   |||||     |||  |
Db        223 GGCTGGGACATCGACTCCCTCTACGACCCGGTGCCCGGGCGCAAGGGCACGACGTACGTC 282

Qy        181 CGAAACGGCGGATTCCTCTACGACGCAGGCCACTTCGACGCCGAATTCTTCGGCATCAGC 240
              || |||| ||  |||||| ||||||| | |    |||||||| |   |||||||| |||
Db        283 CGCAACGCCGCGTTCCTCGACGACGCCGCCGGATTCGACGCGGCCTTCTTCGGGATCTCG 342

Qy        241 CCCCGCGAAGCCCTCGCCATGGACCCCCAGCAACGACTCCTCCTCGAAACCGCCTGGGAA 300
              || ||||| ||||||||||||||||| ||||| || |  |||||||| || |||||||
Db        343 CCGCGCGAGGCCCTCGCCATGGACCCGCAGCAGCGGCAGCTCCTCGAAGCCTCCTGGGAG 402

Qy        301 ACCATCGAACACGCCGGCATCAACCCCCACACCCTCCACGGCACCCCCACCGGAGTCTTC 360
              | |||| |  |||||||| |||||     |  ||| |||||||  |  ||| || | |
Db        403 GTCTTCGAGCGGGCCGGCATCGACCCCGCGTCGGTCCGCGGCACCGACGTCGGCGTGTAC 462

Qy        361 ACCGGCACCAACGGACAGGACTACGCACTTCGCGTGCACAACGCGGGCCAGTCAACCGAT 420
                |||      |  |||||||||||| |    | | |   |||   | |    ||||
Db        463 GTGGGCTGTGGCTACCAGGACTACGCGCCGGACATCCGGGTCGCCCCCGAAGGCACCGGC 522

Qy        421 GGTTTCGCACTGACCGGAACCGCCGGCAGCGTCATCTCCGGTCGTATCTCGTACACGTTT 480
              |||| ||   | |||||  | | ||  |  |||   ||||||  || ||| |||||  |  |
Db        523 GGTTACGTCGTCACCGGCAACTCCTCCGCCGTGGCCTCCGGGCGCATCGCGTACTCCCTC 582

Qy        481 GGTTTTGAGGGTCCTGCGGTGTCGGTGGACACGGCTTGTTCCTCGTCGTTGGTGGCTTTG 540
                 | ||||| || || ||| | |||||||||| || |||||  ||| | || ||  ||
Db        583 GGCCTGGAGGGACCCGCCGTGACCGTGGACACGGCGTGCTCCTCTTCGCTCGTCGCCCTG 642

Qy        541 CATCTGGCCTGTCAGGCGTTGCGTGCGGGTGAGTGCTCGATGGCGCTTGCCGGGGGTGTG 600
              || || |||    |||  ||||    || || ||||||| ||| || |  || || |||
Db        643 CACCTCGCCCTGAAGGGCCTGCGGAACGGCGACTGCTCGACGGCACTCGTGGGCGGCGTG 702

Qy        601 ACGGTGATGTCGTCTCCGGGTGCCTTCGTGGAGTTTTCGCGGCAGCGGGGTCTGGCCGCG 660
              | ||  |  || | ||||| || ||| | |||||      | |||| ||   |||||||
Db        703 GCCGTCCTCGCGACGCCGGGCGCGTTCATCGAGTTCAGCAGCCAGCAGGCCATGGCCGCC 762

Qy        661 GACGGGCATTGCAAGGCGTTCTCGGCGGCGGCGGACGGGACCGGCTGGGGTGAGGGTGTG 720
              |||||  |    |||||  ||| |  |||||||||||||    || |||||| ||||| ||
Db        763 GACGGCCGGACCAAGGGCTTCGCCTCGGCGGCGGACGGCCTCGCCTGGGGCGAGGGCGTC 822

Qy        721 GGGATGCTGCTGGTGGAGCGGCTCTCCGACGCCCATCGCAACGGTCACCGTGTCCTGGCC 780
              |   | || ||  | || ||||||||||||||  |  ||||| || ||||| |||||||||
Db        823 GCCGTACTCCTCCTCGAACGGCTCTCCGACGCGCGGCGCAAGGGCCACCGGGTCCTGGCC 882

Qy        781 GTGGTGCGTGGCAGTGCGGTCAACCAGGACGGTGCGAGCAACGGTCTGACCGCGCCCAAC 840
              || ||||| ||||| ||  |||||||||||||| |||||||||| || || || ||  ||
Db        883 GTCGTGCGCGGCAGCGCCATCAACCAGGACGGCGCGAGCAACGGCCTCACGGCTCCGCAC 942

Qy        841 GGGCCGTCCCAGCAGCGTGTCATCCGCCAGGCCCTCGCCAACGCCGGCTTGTCGGCCGGT 900
              ||||| ||||||||||    | |||||||||||||| ||| ||||  |  |  || | |
Db        943 GGGCCCTCCCAGCAGCACCTGATCCGCCAGGCCCTGGCCGACGCGCGGCTCACGTCGAGC 1002

Qy        901 GATGTCGACGCGGTGGAGGCCCACGGCACCGGCACCACTTTGGGCGACCCGATCGAGGCC 960
              || || ||||  ||||||| |||||||| || |||  | | ||||||||||||||||||
Db       1003 GACGTGGACGTCGTGGAGGGCCACGGCACGGGGACCCGTCTCGGCGACCCGATCGAGGCG 1062

Qy        961 CAGGCCCTCCTCGCGACCTACGGACAGGACCGTGCCGGCGAGGGGCCGCTGTGGCTGGGC 1020
              ||||| || ||||| || ||||| |||| || |||   | |  ||||||| |||||||
Db       1063 CAGGCGCTGCTCGCCACGTACGGGCAGGGGCGCGCCCCGGGGCAGCCGCTGCGGCTGGGG 1122

Qy       1021 TCGGTCAAGTCCAATGTCGGTCACACACAGGCTGCCGCGGGCGTCGCCGGGGTGATCAAG 1080
                || | ||||| ||  |||| |||| ||||| ||  |||| |||||||| || ||||||
Db       1123 ACGCTGAAGTCGAACATCGGGCACACGCAGGCCGCTTCGGGTGTCGCCGGTGTCATCAAG 1182

Qy       1081 ATGGTGATGGCGCTGCGGCATGGTCTGCTGCCGCGGACGTTGCATGTGGATGAGCCGTCG 1140
              ||||||  |||||||| || ||  ||||||||  |||  |||| |||| |||||| || ||
Db       1183 ATGGTGCAGGCGCTGCGCCACGGGGTGCTGCCGAAGACCCTGCACGTGGACGAGCCGACG 1242

Qy       1141 CCGCATGTGGACTGGTCCGCGGGTGCGGTGCAGCTGCTGACGGAGACGGTGCCCTGGCCC 1200
                || || |||||||| || ||| ||||  |||||| || ||| | |||  ||||||
Db       1243 GACCAGGTCGACTGGTCGGCCGGTTCGGTCGAGCTGCTCACCGAGGCCGTGGACTGGCCG 1302

Qy       1201 GGCGGGGAGGGGCGGCTACGGCGGGCAGGAGTGTCATCATTCGGCGTCAGCGGCACCAAC 1260
              |   ||  ||| ||||| || ||||| || || ||  | ||||||||   |||| || |||
Db       1303 GAGCGGCCGGGCCGGCTCCGCCGGGCGGGCGTCTCCGCGTTCGGCGTGGGCGGGACGAAC 1362

Qy       1261 GCCCACGTCATCCTCGAAGAAGCACCCGC 1289
              || |||||| |||| || || || || ||
Db       1363 GCGCACGTCGTCCTGGAGGAGGCCCCGGC 1391


RESULT 15
US-09-428-517-1
; Sequence 1, Application US/09428517
; Patent No. 6251636
; GENERAL INFORMATION:
;  APPLICANT: Betlach, Mary C.
;  APPLICANT: Shah, Sanjay Krishnakant
;  APPLICANT: McDaniel, Robert
;  APPLICANT: Tang, Li
;  TITLE OF INVENTION: RECOMBINANT OLEANDOLIDE POLYKETIDE SYNTHASE
;  FILE REFERENCE: 30062-20029.00
;  CURRENT APPLICATION NUMBER: US/09/428,517
;  CURRENT FILING DATE: 1999-10-28
;  EARLIER APPLICATION NUMBER: 60/120,254
;  EARLIER FILING DATE: 1999-02-16
;  EARLIER APPLICATION NUMBER: 60/106,100
;  EARLIER FILING DATE: 1998-10-29
```

```
;  NUMBER OF SEQ ID NOS: 12
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
;   LENGTH: 50937
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artifical Sequence: Recombinant DNA
US-09-428-517-1

  Query Match           52.2%;  Score 673;  DB 4;  Length 50937;
  Best Local Similarity  70.7%;  Pred. No. 1e-131;
  Matches  911;  Conservative    0;  Mismatches  375;  Indels    3;  Gaps    1;

Qy          1 GAGCCGATCGCCATCGTTGGCATGGCCTGTCGTTTCCCCGGCGGAGTGACCTCGGCGGAC 60
              || ||  ||| |||||| || |||||||| || ||||| |||||  |     || ||||
Db      23046 GACCCCGTCGTCATCGTGGGGATGGCCTGCCGGTTCCCGGGCGGGATCGATACGCCGGAA 23105

Qy         61 GACTTCTGGGATCTGATCTCCTCCGAGCAGGACGCGATCGGCGGATTCCCCACCGACCGC 120
              | ||||||| | ||| ||    | |      || |  |||  || | | ||   || |||||
Db      23106 GCCTTCTGGAAGCTGCTCGAAGCGGGCGGCGATGTCATCTCCGAACTTCCGGCCAACCGC 23165

Qy        121 GGCTGGGACCTGGACACGCTCTACGACCCCGACCCCGACCACCCCGGCACCTGCTACACC 180
              |||||||| ||||     |||     |||| |||||||         |||||| ||  |||
Db      23166 GGCTGGGACATGGAGCGACTCCTGAACCCGGACCCCGAGGCGAAGGGCACCAGCGCCACA 23225

Qy        181 CGAAACGGCGGATTCCTCTACGACGCAGGCCACTTCGACGCCGAATTCTTCGGCATCAGC 240
              ||  ||||||| |||||||||||||| ||  | ||||||||||  |||||||| |||
Db      23226 CGCTACGGCGGTTTCCTCTACGACGCCGGGGAGTTCGACGCCGCCTTCTTCGGTATCTCG 23285

Qy        241 CCCCGCGAAGCCCTCGCCATGGACCCCCAGCAACGACTCCTCCTCGAAACCGCCTGGGAA 300
              || || || ||   | || |||||||| |||||||| || || || ||||||| ||||||
Db      23286 CCGCGTGAGGCGTTGGCGATGGACCCGCAGCAACGGCTGCTGCTGGAAACCGTCTGGGAG 23345

Qy        301 ACCATCGAACACGCCGGCATCAACCCCCACACCCTCCACGGCACCCCCACCGGAGTCTTC 360
                ||||||   ||||||| |    ||| || | |||||| | | ||  |||||    |||
Db      23346 CTCATCGAGAGCGCCGGCGTGGCGCCCGACTCGCTCCACCGGAGCCGGACCGGCACGTTC 23405

Qy        361 ACCGGCACCAACGGACAGGACTACGCACTTCGCGTGCACAACGCGGGCCAGTCAACCGAT 420
              | ||||| |||||| |||  |||||||  |   ||   ||| | |||        ||
Db      23406 ATCGGCAGCAACGGCCAGTTCTACGCACCGCTGCTGTGGAACTCCGGCGGTGATCTGGAG 23465

Qy        421 GGTTTCGCACTGACCGGAACCGCCGGCAGCGTCATCTCCGGTCGTATCTCGTACACGTTT 480
              || | |  |       || | ||||||||||||||| ||||| ||  || | ||| |  |
Db      23466 GGCTACCAAGGCGTGGGCAACGCCGGCAGCGTCATGTCCGGCCGCGTCGCCTACTCCCTC 23525

Qy        481 GGTTTTGAGGGTCCTGCGGTGTCGGTGGACACGGCTTGTTCCTCGTCGTTGGTGGCTTTG 540
              ||| |||||||  |||||||||  |||||||  |||||  |||||  ||||||  |||||||  ||
Db      23526 GGTCTTGAGGGGCCTGCGGTGACGGTGGATACGGCGTGTTCGTCGTCGCTGGTGGCACTG 23585

Qy        541 CATCTGGCCTGTCAGGCGTTGCGTGCGGGTGAGTGCTCGATGGCGCTTGCCGGGGGTGTG 600
              || |||||    |||||| ||||    || ||||||||  | ||  | |||||  ||||||
Db      23586 CACCTGGCGGTGCAGGCGCTGCGCCGTGGCGAGTGCTCACTCGCCATAGCCGGCGGTGTG 23645

Qy        601 ACGGTGATGTCGTCTCCGGGTGCCTTCGTGGAGTTTTCGCGGCAGCGGGGTCTGGCCGCG 660
              ||||||||||   | ||||    ||||||  |||||  || |||||  | |||  ||   ||| |
Db      23646 ACGGTGATGTCCACACCGGACAGCTTCGTTGAGTTCTCACGGCAACAGGGCCTTTCCGAG 23705

Qy        661 GACGGGCATTGCAAGGCGTTCTCGGCGGCGGCGGACGGGACCGGCTGGGGTGAGGGTGTG 720
              |||||  | ||||||||||||| ||     | || || ||   ||||   |  |||||  ||
Db      23706 GACGGCCGTTGCAAGGCGTTCGCGAGCACAGCCGATGGTTTCGGCCTCGCCGAGGGCGTT 23765

Qy        721 GGGATGCTGCTGGTGGAGCGGCTCTCCGACGCCCATCGCAACGGTCACCGTGTCCTGGCC 780
              |  |||| |||||||||||  | || || || |  ||    |||||  || ||  ||||
Db      23766 TCGGCGCTGTTGGTGGAGCGGTTGTCGGATGCGCGGCGGTTGGGTCATCGGGTGTTGGCG 23825

Qy        781 GTGGTGCGTGGCAGTGCGGTCAACCAGGACGGTGCGAGCAACGGTCTGACCGCGCCCAAC 840
              |||||||| || |||||||||||| |||||  ||||||    || ||  ||||  |||||  ||
Db      23826 GTGGTGCGGGGGAGTGCGGTCAATCAGGATGGTGCGTCGAATGGGTTGACGGCGCCGAAT 23885

Qy        841 GGGCCGTCCCAGCAGCGTGTCATCCGCCAGGCCCTCGCCAACGCCGGCTTGTCGGCCGGT 900
              || |||||  |||||||| || || ||   |||  | ||  ||||  ||  || |  | | |
Db      23886 GGTCCGTCGCAGCAGCGGGTGATTCGTGCGGCGTTGGCTGACGCGGGTCTGGCTCCTGCT 23945

Qy        901 GATGTCGACGCGGTGGAGGCCCACGGCACCGGCACCACTTTGGGCGACCCGATCGAGGCC 960
              |||||  ||  | |||||||||  ||  ||  ||  ||  ||     |||||  ||  ||||||||||
Db      23946 GATGTGGATGTGGTGGAGGCGCATGGTACGGGGACGCGGTTGGGTGATCCGATCGAGGCT 24005

Qy        961 CAGGCCCTCCTCGCGACCTACGGACAGGACCGTGCCGGCGAGGGGCCGCTGTGGCTGGGC 1020
              |||||  |  | |||||  ||  ||  ||||  |||||  ||  |  |  |  |||  ||   | ||||
Db      24006 CAGGCGTTGTTGGCGACGTATGGGCAGGGTCGTGCGGGTGGGCGTCCGGTGGTGTTGGGG 24065

Qy       1021 TCGGTCAAGTCCAATGTCGGTCACACACAGGCTGCCGCGGGCGTCGCCGGGGTGATCAAG 1080
              |||||  |||||  ||   ||||  ||  ||  |||||  ||  ||  |||||  ||  ||  |||||  |||
Db      24066 TCGGTGAAGTCGAACATCGGGCATACGCAGGCGGCGGCTGGCGTGGCTGGTGTGATGAAG 24125

Qy       1081 ATGGTGATGGCGCTGCGGCATGGTCTGCTGCCGCGGACGTTGCATGTGGATGAGCCGTCG 1140
              ||||||  ||||||||   ||   |||  ||  |||||   ||||||||||||||||||||||||
Db      24126 ATGGTGCTGGCGCTGGAGCGGGGTGTGGTGCCGAAGACGTTGCATGTGGATGAGCCGTCA 24185

Qy       1141 CCGCATGTGGACTGGTCCGCGGGTGCGGTGCAGCTGCTGACGGAGACGGTGCCCTGGCCC 1200
              ||||||||||||||||| || |||| |||| || ||  |   |||  |||||||  ||| |
Db      24186 CCGCATGTGGACTGGTCGGCTGGTGAGGTGGAGTTGGCGGTTGAGGCGGTGCCGTGGTC- 24244

Qy       1201 GGCGGGGAGGGGCGGCTACGGCGGGCAGGAGTGTCATCATTCGGCGTCAGCGGCACCAAC 1260
                 |||  |||||| | |||||||| || ||||| || ||||  ||||  |||||  ||
Db      24245 --GCGGGGTGGGCGGGTGCGGCGGGCTGGTGTGTCGTCGTTCGGGATCAGTGGCACGAAT 24302

Qy       1261 GCCCACGTCATCCTCGAAGAAGCACCCGC 1289
              || || || ||  | || || || || ||
Db      24303 GCGCATGTGATTGTGGAGGAGGCGCCTGC 24331


Search completed: June 18, 2003, 01:07:22
Job time : 69.8814 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 17, 2003, 17:56:34 ; Search time 152.45 Seconds
(without alignments)
12253.192 Million cell updates/sec

Title: US-09-914-286-1_COPY_6256_7545
Perfect score: 1290
Sequence: 1 gagccgatcgccatcgttgg.....tcctcgaagaagcacccgcc 1290

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1029858 seqs, 724030393 residues

Total number of hits satisfying chosen parameters: 2059716

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
11: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
12: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
13: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
14: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	715.4	55.5	11220	9	US-09-860-846-32	Sequence 32, Appl
2	715.4	55.5	11220	9	US-09-988-384B-32	Sequence 32, Appl
3	715.4	55.5	11220	9	US-09-836-821-32	Sequence 32, Appl
4	715.4	55.5	11220	10	US-09-861-289-32	Sequence 32, Appl
5	715.4	55.5	36778	9	US-09-860-846-5	Sequence 5, Appli
6	715.4	55.5	36778	9	US-09-836-821-5	Sequence 5, Appli
7	715.4	55.5	36778	10	US-09-861-289-5	Sequence 5, Appli
8	715.4	55.5	37948	9	US-09-988-384B-5	Sequence 5, Appli
9	715.4	55.5	38506	9	US-09-793-708-19	Sequence 19, Appl
10	689.4	53.4	13842	9	US-09-860-846-30	Sequence 30, Appl
11	689.4	53.4	13842	9	US-09-988-384B-30	Sequence 30, Appl
12	689.4	53.4	13842	9	US-09-836-821-30	Sequence 30, Appl
13	689.4	53.4	13842	10	US-09-861-289-30	Sequence 30, Appl
14	680.4	52.7	15872	9	US-09-860-846-1	Sequence 1, Appli
15	680.4	52.7	15872	9	US-09-988-384B-1	Sequence 1, Appli
16	680.4	52.7	15872	9	US-09-836-821-1	Sequence 1, Appli
17	680.4	52.7	15872	10	US-09-861-289-1	Sequence 1, Appli
18	673	52.2	4041	9	US-09-860-846-36	Sequence 36, Appl
19	673	52.2	4041	9	US-09-988-384B-36	Sequence 36, Appl
20	673	52.2	4041	9	US-09-836-821-36	Sequence 36, Appl
21	673	52.2	4041	10	US-09-861-289-36	Sequence 36, Appl
22	673	52.2	50937	9	US-09-808-880-1	Sequence 1, Appli
23	655.2	50.8	4689	9	US-09-860-846-34	Sequence 34, Appl
24	655.2	50.8	4689	9	US-09-988-384B-34	Sequence 34, Appl
25	655.2	50.8	4689	9	US-09-836-821-34	Sequence 34, Appl
26	655.2	50.8	4689	10	US-09-861-289-34	Sequence 34, Appl
27	522.6	40.5	68750	9	US-10-014-717-1	Sequence 1, Appli
28	387.4	30.0	5484	9	US-09-712-363-115	Sequence 115, App
29	262.6	20.4	4209	9	US-09-712-363-20	Sequence 20, Appl
30	239.4	18.6	4830	9	US-09-738-626-3166	Sequence 3166, Ap
c 31	239.4	18.6	3309400	9	US-09-738-626-1	Sequence 1, Appli
32	216	16.7	8460	9	US-10-237-271-2	Sequence 2, Appli
33	214.2	16.6	615	9	US-10-125-815-4	Sequence 4, Appli
34	209.4	16.2	615	9	US-10-125-815-3	Sequence 3, Appli
35	184.8	14.3	873	10	US-09-833-381-1909	Sequence 1909, Ap
36	166.4	12.9	7692	9	US-09-836-705-45	Sequence 45, Appl
37	165.4	12.8	9099	9	US-09-836-705-43	Sequence 43, Appl
38	162.2	12.6	6000	10	US-09-775-938A-29	Sequence 29, Appl
39	141.4	11.0	8730	9	US-10-124-800-1	Sequence 1, Appli
40	140.2	10.9	643	10	US-09-924-256A-53	Sequence 53, Appl
41	114.8	8.9	655	10	US-09-924-256A-45	Sequence 45, Appl
42	111.6	8.7	643	10	US-09-924-256A-31	Sequence 31, Appl
43	108.2	8.4	722	9	US-09-836-705-17	Sequence 17, Appl
44	105.6	8.2	637	10	US-09-924-256A-21	Sequence 21, Appl
45	104.6	8.1	8268	9	US-10-331-061-83	Sequence 83, Appl

ALIGNMENTS

RESULT 1
US-09-860-846-32
; Sequence 32, Application US/09860846
; Patent No. US20020164742A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/860,846
; CURRENT FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 32
; LENGTH: 11220
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-860-846-32

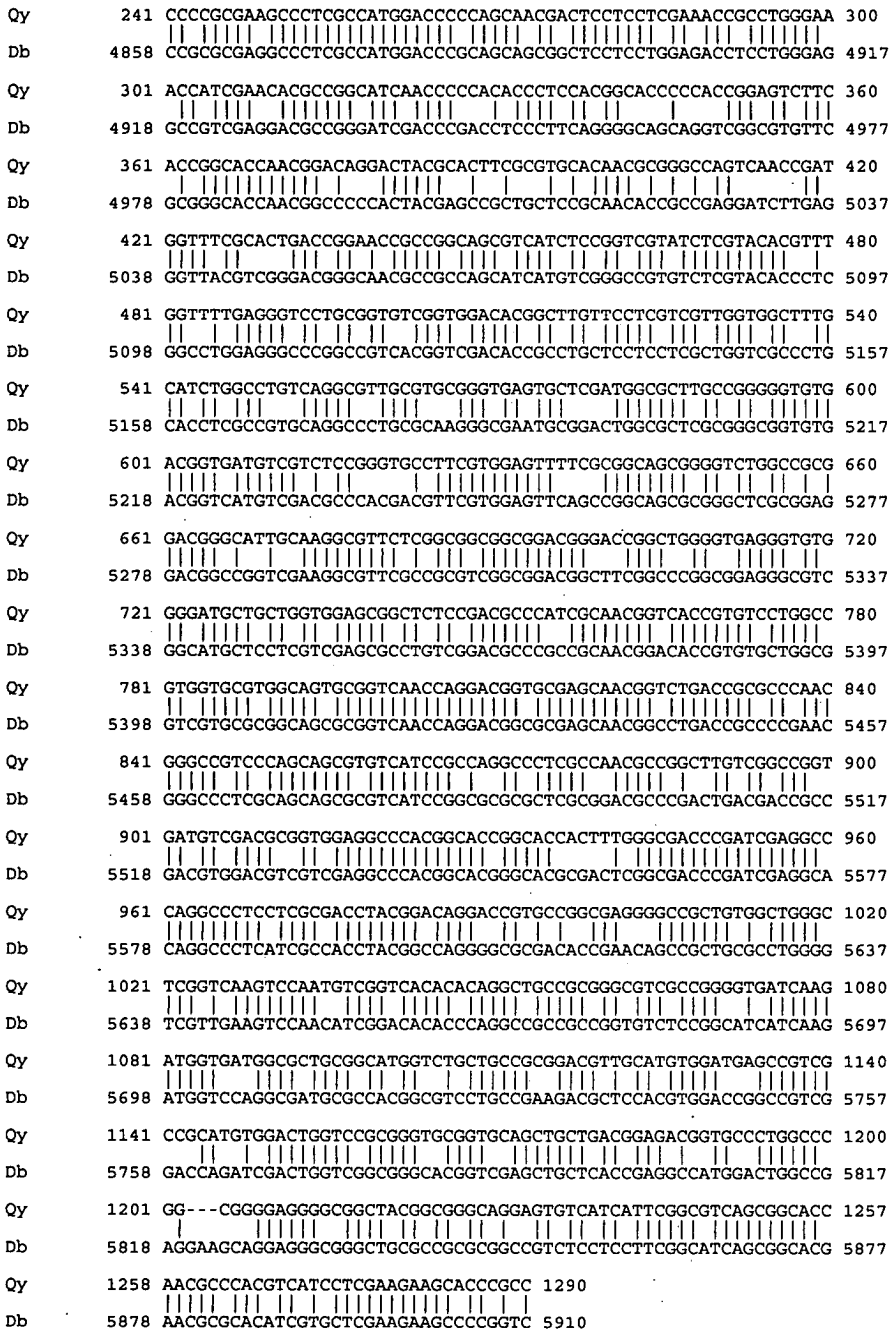
Query Match 55.5%; Score 715.4; DB 9; Length 11220;
Best Local Similarity 72.6%; Pred. No. 8.4e-165;
Matches 939; Conservative 0; Mismatches 351; Indels 3; Gaps 1;

```
Qy          1 GAGCCGATCGCCATCGTTGGCATGGCCTGTCGTTTCCCCGGCGGAGTGACCTCGGCGGAC 60
              ||||| |||| ||||| |||||||| || ||  | || ||||| ||  ||||| ||||
Db       4618 GAGCCCGTCGCGATCGTCGGCATGGCGTGCCGCCTGCCGGGCGGTGTCGCCTCGCCGGAG 4677

Qy         61 GACTTCTGGGATCTGATCTCCTCCGAGCAGGACGCGATCGGCGGATTCCCCACCGACCGC 120
              ||| | |||   ||| |  ||  ||   ||||||||||  ||| |||||     ||||||
Db       4678 GACCTGTGGCGGCTGGTGGCCGGCGGCGAGGACGCGATCTCCGGCTTCCCGCAGGACCGC 4737

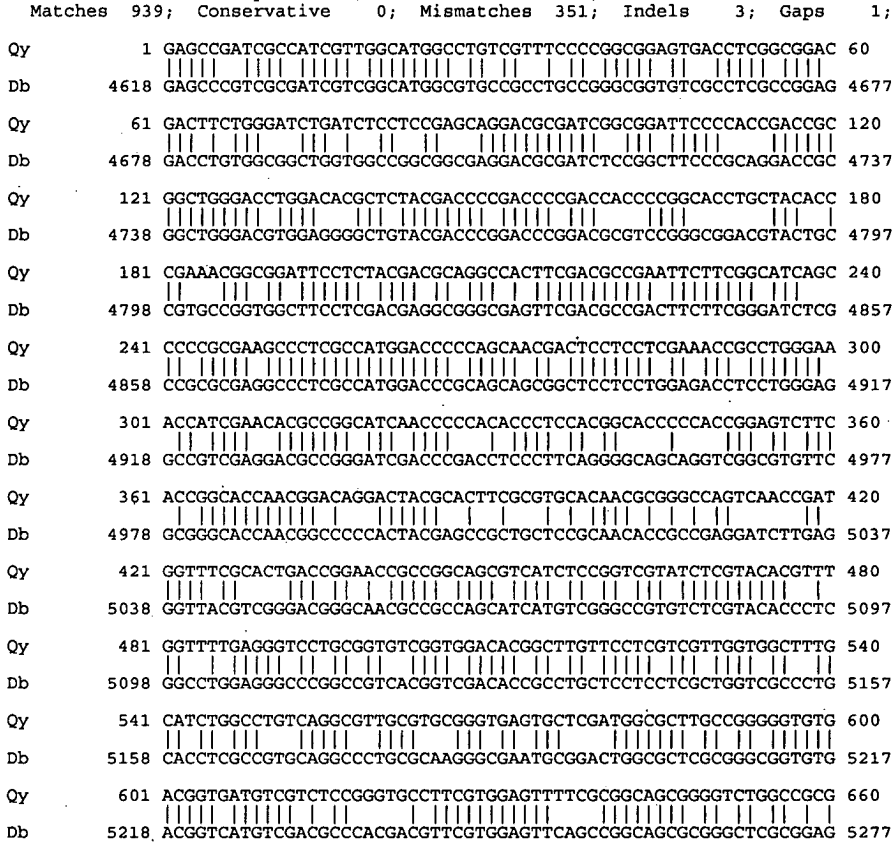
Qy        121 GGCTGGGACCTGGACACGCTCTACGACCCCGACCCCGACCACCCCGGCACCTGCTACACC 180
              |||||||| ||||   ||| |||||||| ||||| |||    ||||        |||  |
Db       4738 GGCTGGGACGTGGAGGGGCTGTACGACCCGGACCCGGACGCGTCCGGGCGGACGTACTGC 4797

Qy        181 CGAAACGGCGGATTCCTCTACGACGCAGGCCACTTCGACGCCGAATTCTTCGGCATCAGC 240
              ||   ||| || |||||| |||| || ||| | |||||||||||| ||||||||  |||
Db       4798 CGTGCCGGTGGCTTCCTCGACGAGGCGGGCGAGTTCGACGCCGACTTCTTCGGGATCTCG 4857
```



RESULT 2
US-09-988-384B-32
; Sequence 32, Application US/09988384B
; Publication No. US20030073824A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.536US1
; CURRENT APPLICATION NUMBER: US/09/988,384B
; CURRENT FILING DATE: 2001-11-19
; PRIOR APPLICATION NUMBER: PCT/US99/14398
; PRIOR FILING DATE: 1999-06-25
; PRIOR APPLICATION NUMBER: US 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 53
; SEQ ID NO 32
; LENGTH: 11220
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-988-384B-32

Query Match 55.5%; Score 715.4; DB 9; Length 11220;
Best Local Similarity 72.6%; Pred. No. 8.4e-165;
Matches 939; Conservative 0; Mismatches 351; Indels 3; Gaps 1;



```
Qy       661 GACGGGCATTGCAAGGCGTTCTCGGCGGCGGCGGACGGGACCGGCTGGGGTGAGGGTGTG 720
Db      5278 GACGGCCGGTCGAAGGCGTTCGCCGCGTCGGCGGACGGCTTCGGCCCGGCGGAGGGCGTC 5337

Qy       721 GGGATGCTGCTGGTGGAGCGGCTCTCCGACGCCCATCGCAACGGTCACCGTGTCCTGGCC 780
Db      5338 GGCATGCTCCTCGTCGAGCGCCTGTCGGACGCCCGCCGCAACGGACACCGTGTGCTGGCG 5397

Qy       781 GTGGTGCGTGGCAGTGCGGTCAACCAGGACGGTGCGAGCAACGGTCTGACCGCGCCCAAC 840
Db      5398 GTCGTGCGCGGCAGCGCGGTCAACCAGGACGGCGCGAGCAACGGCCTGACCGCCCCGAAC 5457

Qy       841 GGGCCGTCCCAGCAGCGTGTCATCCGCCAGGCCCTCGCCAACGCCGGCTTGTCGGCCGGT 900
Db      5458 GGGCCCTCGCAGCAGCGCGTCATCCGGCGCGCGCTCGCGGACGCCCGACTGACGACCGCC 5517

Qy       901 GATGTCGACGCGGTGGAGGCCCACGGCACCGGCACCACTTTGGGCGACCCGATCGAGGCC 960
Db      5518 GACGTGGACGTCGTCGAGGCCCACGGCACGGGCACGCGACTCGGCGACCCGATCGAGGCA 5577

Qy       961 CAGGCCCTCCTCGCGACCTACGGACAGGACCGTGCCGGCGAGGGGCCGCTGTGGCTGGGC 1020
Db      5578 CAGGCCCTCATCGCCACCTACGGCCAGGGGCGCGACACCGAACAGCCGCTGCGCCTGGGG 5637

Qy      1021 TCGGTCAAGTCCAATGTCGGTCACACACAGGCTGCCGCGGGCGTCGCCGGGGTGATCAAG 1080
Db      5638 TCGTTGAAGTCCAACATCGGACACACCCAGGCCGCCGCCGGTGTCTCCGGCATCATCAAG 5697

Qy      1081 ATGGTGATGGCGCTGCGGCATGGTCTGCTGCCGCGGACGTTGCATGTGGATGAGCCGTCG 1140
Db      5698 ATGGTCCAGGCGATGCGCCACGGCGTCCTGCCGAAGACGCTCCACGTGGACCGGCCGTCG 5757

Qy      1141 CCGCATGTGGACTGGTCCGCGGGTGCGGTGCAGCTGCTGACGGAGACGGTGCCCTGGCCC 1200
Db      5758 GACCAGATCGACTGGTCGGCGGGCACGGTCGAGCTGCTCACCGAGGCCATGGACTGGCCG 5817

Qy      1201 GG---CGGGGAGGGGCGGCTACGGCGGGCAGGAGTGTCATCATTCGGCGTCAGCGGCACC 1257
Db      5818 AGGAAGCAGGAGGGCGGGCTGCGCCGCGCGGCCGTCTCCTCCTTCGGCATCAGCGGCACG 5877

Qy      1258 AACGCCCACGTCATCCTCGAAGAAGCACCCGCC 1290
Db      5878 AACGCGCACATCGTGCTCGAAGAAGCCCCGGTC 5910
```

```
RESULT 3
US-09-836-821-32
; Sequence 32, Application US/09836821
; Publication No. US20030087405A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/836,821
;  CURRENT FILING DATE:  2001-04-17
;  PRIOR APPLICATION NUMBER: 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 32
;   LENGTH: 11220
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-836-821-32

  Query Match             55.5%;  Score 715.4;  DB 9;  Length 11220;
  Best Local Similarity   72.6%;  Pred. No. 8.4e-165;
  Matches  939;  Conservative    0;  Mismatches  351;  Indels    3;  Gaps    1;

Qy          1 GAGCCGATCGCCATCGTTGGCATGGCCTGTCGTTTCCCCGGCGGAGTGACCTCGGCGGAC 60
```

```
Db      4618 GAGCCCGTCGCGATCGTCGGCATGGCGTGCCGCCTGCCGGGCGGTGTCGCCTCGCCGGAG 4677

Qy        61 GACTTCTGGGATCTGATCTCCTCCGAGCAGGACGCGATCGGCGGATTCCCCACCGACCGC 120
Db      4678 GACCTGTGGCGGCTGGTGGCCGGCGGCGAGGACGCGATCTCCGGCTTCCCGCAGGACCGC 4737

Qy       121 GGCTGGGACCTGGACACGCTCTACGACCCCGACCCCGACCACCCCGGCACCTGCTACACC 180
Db      4738 GGCTGGGACGTGGAGGGGCTGTACGACCCGGACCCGGACGCGTCCGGGCGGACGTACTGC 4797

Qy       181 CGAAACGGCGGATTCCTCTACGACGCAGGCCACTTCGACGCCGAATTCTTCGGCATCAGC 240
Db      4798 CGTGCCGGTGGCTTCCTCGACGAGGCGGGCGAGTTCGACGCCGACTTCTTCGGGATCTCG 4857

Qy       241 CCCCGCGAAGCCCTCGCCATGGACCCCCAGCAACGACTCCTCCTCGAAACCGCCTGGGAA 300
Db      4858 CCGCGCGAGGCCCTCGCCATGGACCCGCAGCAGCGGCTCCTCCTGGAGACCTCCTGGGAG 4917

Qy       301 ACCATCGAACACGCCGGCATCAACCCCCACACCCTCCACGGCACCCCCACCGGAGTCTTC 360
Db      4918 GCCGTCGAGGACGCCGGGATCGACCCGACCTCCCTTCAGGGGCAGCAGGTCGGCGTGTTC 4977

Qy       361 ACCGGCACCAACGGACAGGACTACGCACTTCGCGTGCACAACGCGGGCCAGTCAACCGAT 420
Db      4978 GCGGGCACCAACGGCCCCCACTACGAGCCGCTGCTCCGCAACACCGCCGAGGATCTTGAG 5037

Qy       421 GGTTTCGCACTGACCGGAACCGCCGGCAGCGTCATCTCCGGTCGTATCTCGTACACGTTT 480
Db      5038 GGTTACGTCGGGACGGGCAACGCCGCCAGCATCATGTCGGGCCGTGTCTCGTACACCCTC 5097

Qy       481 GGTTTTGAGGGTCCTGCGGTGTCGGTGGACACGGCTTGTTCCTCGTCGTTGGTGGCTTTG 540
Db      5098 GGCCTGGAGGGCCCGGCCGTCACGGTCGACACCGCCTGCTCCTCCTCGCTGGTCGCCCTG 5157

Qy       541 CATCTGGCCTGTCAGGCGTTGCGTGCGGGTGAGTGCTCGATGGCGCTTGCCGGGGGTGTG 600
Db      5158 CACCTCGCCGTGCAGGCCCTGCGCAAGGGCGAATGCGGACTGGCGCTCGCGGGCGGTGTG 5217

Qy       601 ACGGTGATGTCGTCTCCGGGTGCCTTCGTGGAGTTTTCGCGGCAGCGGGGTCTGGCCGCG 660
Db      5218 ACGGTCATGTCGACGCCCACGACGTTCGTGGAGTTCAGCCGGCAGCGCGGGCTCGCGGAG 5277

Qy       661 GACGGGCATTGCAAGGCGTTCTCGGCGGCGGCGGACGGGACCGGCTGGGGTGAGGGTGTG 720
Db      5278 GACGGCCGGTCGAAGGCGTTCGCCGCGTCGGCGGACGGCTTCGGCCCGGCGGAGGGCGTC 5337

Qy       721 GGGATGCTGCTGGTGGAGCGGCTCTCCGACGCCCATCGCAACGGTCACCGTGTCCTGGCC 780
Db      5338 GGCATGCTCCTCGTCGAGCGCCTGTCGGACGCCCGCCGCAACGGACACCGTGTGCTGGCG 5397

Qy       781 GTGGTGCGTGGCAGTGCGGTCAACCAGGACGGTGCGAGCAACGGTCTGACCGCGCCCAAC 840
Db      5398 GTCGTGCGCGGCAGCGCGGTCAACCAGGACGGCGCGAGCAACGGCCTGACCGCCCCGAAC 5457

Qy       841 GGGCCGTCCCAGCAGCGTGTCATCCGCCAGGCCCTCGCCAACGCCGGCTTGTCGGCCGGT 900
Db      5458 GGGCCCTCGCAGCAGCGCGTCATCCGGCGCGCGCTCGCGGACGCCCGACTGACGACCGCC 5517

Qy       901 GATGTCGACGCGGTGGAGGCCCACGGCACCGGCACCACTTTGGGCGACCCGATCGAGGCC 960
Db      5518 GACGTGGACGTCGTCGAGGCCCACGGCACGGGCACGCGACTCGGCGACCCGATCGAGGCA 5577

Qy       961 CAGGCCCTCCTCGCGACCTACGGACAGGACCGTGCCGGCGAGGGGCCGCTGTGGCTGGGC 1020
Db      5578 CAGGCCCTCATCGCCACCTACGGCCAGGGGCGCGACACCGAACAGCCGCTGCGCCTGGGG 5637

Qy      1021 TCGGTCAAGTCCAATGTCGGTCACACACAGGCTGCCGCGGGCGTCGCCGGGGTGATCAAG 1080
Db      5638 TCGTTGAAGTCCAACATCGGACACACCCAGGCCGCCGCCGGTGTCTCCGGCATCATCAAG 5697

Qy      1081 ATGGTGATGGCGCTGCGGCATGGTCTGCTGCCGCGGACGTTGCATGTGGATGAGCCGTCG 1140
```

```
Db       5698 ATGGTCCAGGCGATGCGCCACGGCGTCCTGCCGAAGACGCTCCACGTGGACCGGCCGTCG 5757

Qy       1141 CCGCATGTGGACTGGTCCGCGGGTGCGGTGCAGCTGCTGACGGAGACGGTGCCCTGGCCC 1200

Db       5758 GACCAGATCGACTGGTCGGCGGGCACGGTCGAGCTGCTCACCGAGGCCATGGACTGGCCG 5817

Qy       1201 GG---CGGGGAGGGGCGGCTACGGCGGGCAGGAGTGTCATCATTCGGCGTCAGCGGCACC 1257

Db       5818 AGGAAGCAGGAGGGCGGGCTGCGCCGCGCGGCCGTCTCCTCCTTCGGCATCAGCGGCACG 5877

Qy       1258 AACGCCCACGTCATCCTCGAAGAAGCACCCGCC 1290

Db       5878 AACGCGCACATCGTGCTCGAAGAAGCCCCGGTC 5910


RESULT 4
US-09-861-289-32
; Sequence 32, Application US/09861289
; Patent No. US20020110897A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/861,289
;  CURRENT FILING DATE:  2001-05-18
;  PRIOR APPLICATION NUMBER: 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 32
;   LENGTH: 11220
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-861-289-32

  Query Match             55.5%;  Score 715.4;  DB 10;  Length 11220;
  Best Local Similarity   72.6%;  Pred. No. 8.4e-165;
  Matches  939;  Conservative    0;  Mismatches  351;  Indels    3;  Gaps    1;

Qy          1 GAGCCGATCGCCATCGTTGGCATGGCCTGTCGTTTCCCCGGCGGAGTGACCTCGGCGGAC 60

Db       4618 GAGCCCGTCGCGATCGTCGGCATGGCGTGCCGCCTGCCGGGCGGTGTCGCCTCGCCGGAG 4677

Qy         61 GACTTCTGGGATCTGATCTCCTCCGAGCAGGACGCGATCGGCGGATTCCCCACCGACCGC 120

Db       4678 GACCTGTGGCGGCTGGTGGCCGGCGGCGAGGACGCGATCTCCGGCTTCCCGCAGGACCGC 4737

Qy         121 GGCTGGGACCTGGACACGCTCTACGACCCCGACCCCGACCACCCCGGCACCTGCTACACC 180

Db       4738 GGCTGGGACGTGGAGGGGCTGTACGACCCGGACCCGGACGCGTCCGGGCGGACGTACTGC 4797

Qy         181 CGAAACGGCGGATTCCTCTACGACGCAGGCCACTTCGACGCCGAATTCTTCGGCATCAGC 240

Db       4798 CGTGCCGGTGGCTTCCTCGACGAGGCGGGCGAGTTCGACGCCGACTTCTTCGGGATCTCG 4857

Qy         241 CCCCGCGAAGCCCTCGCCATGGACCCCCAGCAACGACTCCTCCTCGAAACCGCCTGGGAA 300

Db       4858 CCGCGCGAGGCCCTCGCCATGGACCCGCAGCAGCGGCTCCTCCTGGAGACCTCCTGGGAG 4917

Qy         301 ACCATCGAACACGCCGGCATCAACCCCCACACCCTCCACGGCACCCCCACCGGAGTCTTC 360

Db       4918 GCCGTCGAGGACGCCGGGATCGACCCGACCTCCCTTCAGGGGCAGCAGGTCGGCGTGTTC 4977

Qy         361 ACCGGCACCAACGGACAGGACTACGCACTTCGCGTGCACAACGCGGGCCAGTCAACCGAT 420

Db       4978 GCGGGCACCAACGGCCCCCACTACGAGCCGCTGCTCCGCAACACCGCCGAGGATCTTGAG 5037

Qy         421 GGTTTCGCACTGACCGGAACCGCCGGCAGCGTCATCTCCGGTCGTATCTCGTACACGTTT 480

Db       5038 GGTTACGTCGGGACGGGCAACGCCGCCAGCATCATGTCGGGCCGTGTCTCGTACACCCTC 5097

Qy         481 GGTTTTGAGGGTCCTGCGGTGTCGGTGGACACGGCTTGTTCCTCGTCGTTGGTGGCTTTG 540

Db       5098 GGCCTGGAGGGCCCGGCCGTCACGGTCGACACCGCCTGCTCCTCCTCGCTGGTCGCCCTG 5157

Qy         541 CATCTGGCCTGTCAGGCGTTGCGTGCGGGTGAGTGCTCGATGGCGCTTGCCGGGGGTGTG 600

Db       5158 CACCTCGCCGTGCAGGCCCTGCGCAAGGGCGAATGCGGACTGGCGCTCGCGGGCGGTGTG 5217

Qy         601 ACGGTGATGTCGTCTCCGGGTGCCTTCGTGGAGTTTTCGCGGCAGCGGGGTCTGGCCGCG 660

Db       5218 ACGGTCATGTCGACGCCCACGACGTTCGTGGAGTTCAGCCGGCAGCGCGGGCTCGCGGAG 5277

Qy         661 GACGGGCATTGCAAGGCGTTCTCGGCGGCGGCGGACGGGACCGGCTGGGGTGAGGGTGTG 720

Db       5278 GACGGCCGGTCGAAGGCGTTCGCCGCGTCGGCGGACGGCTTCGGCCCGGCGGAGGGCGTC 5337

Qy         721 GGGATGCTGCTGGTGGAGCGGCTCTCCGACGCCCATCGCAACGGTCACCGTGTCCTGGCC 780

Db       5338 GGCATGCTCCTCGTCGAGCGCCTGTCGGACGCCCGCCGCAACGGACACCGTGTGCTGGCG 5397

Qy         781 GTGGTGCGTGGCAGTGCGGTCAACCAGGACGGTGCGAGCAACGGTCTGACCGCGCCCAAC 840

Db       5398 GTCGTGCGCGGCAGCGCGGTCAACCAGGACGGCGCGAGCAACGGCCTGACCGCCCCGAAC 5457

Qy         841 GGGCCGTCCCAGCAGCGTGTCATCCGCCAGGCCCTCGCCAACGCCGGCTTGTCGGCCGGT 900

Db       5458 GGGCCCTCGCAGCAGCGCGTCATCCGGCGCGCGCTCGCGGACGCCCGACTGACGACCGCC 5517

Qy         901 GATGTCGACGCGGTGGAGGCCCACGGCACCGGCACCACTTTGGGCGACCCGATCGAGGCC 960

Db       5518 GACGTGGACGTCGTCGAGGCCCACGGCACGGGCACGCGACTCGGCGACCCGATCGAGGCA 5577

Qy         961 CAGGCCCTCCTCGCGACCTACGGACAGGACCGTGCCGGCGAGGGGCCGCTGTGGCTGGGC 1020

Db       5578 CAGGCCCTCATCGCCACCTACGGCCAGGGGCGCGACACCGAACAGCCGCTGCGCCTGGGG 5637

Qy        1021 TCGGTCAAGTCCAATGTCGGTCACACACAGGCTGCCGCGGGCGTCGCCGGGGTGATCAAG 1080

Db       5638 TCGTTGAAGTCCAACATCGGACACACCCAGGCCGCCGCCGGTGTCTCCGGCATCATCAAG 5697

Qy        1081 ATGGTGATGGCGCTGCGGCATGGTCTGCTGCCGCGGACGTTGCATGTGGATGAGCCGTCG 1140

Db       5698 ATGGTCCAGGCGATGCGCCACGGCGTCCTGCCGAAGACGCTCCACGTGGACCGGCCGTCG 5757

Qy        1141 CCGCATGTGGACTGGTCCGCGGGTGCGGTGCAGCTGCTGACGGAGACGGTGCCCTGGCCC 1200

Db       5758 GACCAGATCGACTGGTCGGCGGGCACGGTCGAGCTGCTCACCGAGGCCATGGACTGGCCG 5817

Qy        1201 GG---CGGGGAGGGGCGGCTACGGCGGGCAGGAGTGTCATCATTCGGCGTCAGCGGCACC 1257

Db       5818 AGGAAGCAGGAGGGCGGGCTGCGCCGCGCGGCCGTCTCCTCCTTCGGCATCAGCGGCACG 5877

Qy        1258 AACGCCCACGTCATCCTCGAAGAAGCACCCGCC 1290

Db       5878 AACGCGCACATCGTGCTCGAAGAAGCCCCGGTC 5910


RESULT 5
US-09-860-846-5
; Sequence 5, Application US/09860846
; Patent No. US20020164742A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/860,846
;  CURRENT FILING DATE:  2001-05-18
;  PRIOR APPLICATION NUMBER: 09/105,537
;  PRIOR FILING DATE: 1998-06-26
```

```
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 5
;   LENGTH: 36778
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-860-846-5

  Query Match           55.5%;  Score 715.4;  DB 9;  Length 36778;
  Best Local Similarity  72.6%;  Pred. No. 8.1e-165;
  Matches  939;  Conservative    0;  Mismatches  351;  Indels    3;  Gaps    1;

Qy          1 GAGCCGATCGCCATCGTTGGCATGGCCTGTCGTTTCCCCGGCGGAGTGACCTCGGCGGAC 60
              ||||| |||| ||||| |||||||| || ||  | || ||||| ||  ||||| ||||
Db      20305 GAGCCCGTCGCGATCGTCGGCATGGCGTGCCGCCTGCCGGGCGGTGTCGCCTCGCCGGAG 20364

Qy         61 GACTTCTGGGATCTGATCTCCTCCGAGCAGGACGCGATCGGCGGATTCCCCACCGACCGC 120
              ||| | |||   ||| |  ||  ||   |||||||||||  ||| |||||    ||||||
Db      20365 GACCTGTGGCGGCTGGTGGCCGGCGGCGAGGACGCGATCTCCGGCTTCCCGCAGGACCGC 20424

Qy        121 GGCTGGGACCTGGACACGCTCTACGACCCCGACCCCGACCACCCCGGCACCTGCTACACC 180
              |||||||| ||||   ||| |||||||| ||||| |||     |||          |||  |
Db      20425 GGCTGGGACGTGGAGGGGCTGTACGACCCGGACCCGGACGCGTCCGGGCGGACGTACTGC 20484

Qy        181 CGAAACGGCGGATTCCTCTACGACGCAGGCCACTTCGACGCCGAATTCTTCGGCATCAGC 240
              ||   ||| || |||||| |||| || ||| | ||||||||||| ||||||||  |||
Db      20485 CGTGCCGGTGGCTTCCTCGACGAGGCGGGCGAGTTCGACGCCGACTTCTTCGGGATCTCG 20544

Qy        241 CCCCGCGAAGCCCTCGCCATGGACCCCCAGCAACGACTCCTCCTCGAAACCGCCTGGGAA 300
              || ||||| |||||||||||||||||| ||||| || |||||||| || ||| |||||||
Db      20545 CCGCGCGAGGCCCTCGCCATGGACCCGCAGCAGCGGCTCCTCCTGGAGACCTCCTGGGAG 20604

Qy        301 ACCATCGAACACGCCGGCATCAACCCCCACACCCTCCACGGCACCCCCACCGGAGTCTTC 360
              || ||||  ||||||| ||| ||||    | |||| || ||     |    ||| || |||
Db      20605 GCCGTCGAGGACGCCGGGATCGACCCGACCTCCCTTCAGGGGCAGCAGGTCGGCGTGTTC 20664

Qy        361 ACCGGCACCAACGGACAGGACTACGCACTTCGCGTGCACAACGCGGGCCAGTCAACCGAT 420
              | ||||||||||| |   ||||||  |   |   | | |||| | | | ||      ||
Db      20665 GCGGGCACCAACGGCCCCCACTACGAGCCGCTGCTCCGCAACACCGCCGAGGATCTTGAG 20724

Qy        421 GGTTTCGCACTGACCGGAACCGCCGGCAGCGTCATCTCCGGTCGTATCTCGTACACGTTT 480
              |||| ||    ||| || | ||||| |||| |||| || || ||| ||||||||||  |
Db      20725 GGTTACGTCGGGACGGGCAACGCCGCCAGCATCATGTCGGGCCGTGTCTCGTACACCCTC 20784

Qy        481 GGTTTTGAGGGTCCTGCGGTGTCGGTGGACACGGCTTGTTCCTCGTCGTTGGTGGCTTTG 540
              ||  | ||||| || || ||  |||| |||||  || || ||||| ||| |||| ||  ||
Db      20785 GGCCTGGAGGGCCCGGCCGTCACGGTCGACACCGCCTGCTCCTCCTCGCTGGTCGCCCTG 20844

Qy        541 CATCTGGCCTGTCAGGCGTTGCGTGCGGGTGAGTGCTCGATGGCGCTTGCCGGGGGTGTG 600
              || || |||   |||||  ||||   ||| || |||    |||||| || || ||||||
Db      20845 CACCTCGCCGTGCAGGCCCTGCGCAAGGGCGAATGCGGACTGGCGCTCGCGGGCGGTGTG 20904

Qy        601 ACGGTGATGTCGTCTCCGGGTGCCTTCGTGGAGTTTTCGCGGCAGCGGGGTCTGGCCGCG 660
              ||||| |||||| | ||      | ||||||||||    |||||||| || || || | |
Db      20905 ACGGTCATGTCGACGCCCACGACGTTCGTGGAGTTCAGCCGGCAGCGCGGGCTCGCGGAG 20964

Qy        661 GACGGGCATTGCAAGGCGTTCTCGGCGGCGGCGGACGGGACCGGCTGGGGTGAGGGTGTG 720
              ||||| |  | ||||||||| | ||| ||||||||||   ||||  ||  ||||| ||
Db      20965 GACGGCCGGTCGAAGGCGTTCGCCGCGTCGGCGGACGGCTTCGGCCCGGCGGAGGGCGTC 21024

Qy        721 GGGATGCTGCTGGTGGAGCGGCTCTCCGACGCCCATCGCAACGGTCACCGTGTCCTGGCC 780
              || ||||| || || ||||| || || |||||||  |||||||| |||||||| |||||
Db      21025 GGCATGCTCCTCGTCGAGCGCCTGTCGGACGCCCGCCGCAACGGACACCGTGTGCTGGCG 21084

Qy        781 GTGGTGCGTGGCAGTGCGGTCAACCAGGACGGTGCGAGCAACGGTCTGACCGCGCCCAAC 840
              || ||||| ||||| ||||||||||||||||| |||||||||| |||||||| || |||
Db      21085 GTCGTGCGCGGCAGCGCGGTCAACCAGGACGGCGCGAGCAACGGCCTGACCGCCCCGAAC 21144

Qy        841 GGGCCGTCCCAGCAGCGTGTCATCCGCCAGGCCCTCGCCAACGCCGGCTTGTCGGCCGGT 900
              ||||| || |||||||| |||||||| |  || |||||  ||||| |  || || |||
Db      21145 GGGCCCTCGCAGCAGCGCGTCATCCGGCGCGCGCTCGCGGACGCCCGACTGACGACCGCC 21204

Qy        901 GATGTCGACGCGGTGGAGGCCCACGGCACCGGCACCACTTTGGGCGACCCGATCGAGGCC 960
              || || ||||  || ||||||||||||||| |||||     | ||||||||||||||||||
Db      21205 GACGTGGACGTCGTCGAGGCCCACGGCACGGGCACGCGACTCGGCGACCCGATCGAGGCA 21264

Qy        961 CAGGCCCTCCTCGCGACCTACGGACAGGACCGTGCCGGCGAGGGGCCGCTGTGGCTGGGC 1020
              ||||||||| |||| |||||||| ||||  || | |  |||   ||||||| | |||||
Db      21265 CAGGCCCTCATCGCCACCTACGGCCAGGGGCGCGACACCGAACAGCCGCTGCGCCTGGGG 21324

Qy       1021 TCGGTCAAGTCCAATGTCGGTCACACACAGGCTGCCGCGGGCGTCGCCGGGGTGATCAAG 1080
              ||| | ||||||||  |||| |||||  |||||  |||| || ||| ||||  | ||||||
Db      21325 TCGTTGAAGTCCAACATCGGACACACCCAGGCCGCCGCCGGTGTCTCCGGCATCATCAAG 21384

Qy       1081 ATGGTGATGGCGCTGCGGCATGGTCTGCTGCCGCGGACGTTGCATGTGGATGAGCCGTCG 1140
              |||||  |||| |||| || ||  | ||||||  |||| | || |||||   |||||||
Db      21385 ATGGTCCAGGCGATGCGCCACGGCGTCCTGCCGAAGACGCTCCACGTGGACCGGCCGTCG 21444

Qy       1141 CCGCATGTGGACTGGTCCGCGGGTGCGGTGCAGCTGCTGACGGAGACGGTGCCCTGGCCC 1200
                ||  | |||||||| |||||  ||||  ||||||| || ||| |  ||  ||||||
Db      21445 GACCAGATCGACTGGTCGGCGGGCACGGTCGAGCTGCTCACCGAGGCCATGGACTGGCCG 21504

Qy       1201 GG---CGGGGAGGGGCGGCTACGGCGGGCAGGAGTGTCATCATTCGGCGTCAGCGGCACC 1257
              |      ||||||  |||| || || || |  || || || |||||| ||||||||||
Db      21505 AGGAAGCAGGAGGGCGGGCTGCGCCGCGCGGCCGTCTCCTCCTTCGGCATCAGCGGCACG 21564

Qy       1258 AACGCCCACGTCATCCTCGAAGAAGCACCCGCC 1290
              ||||| ||| || | |||||||||| || | |
Db      21565 AACGCGCACATCGTGCTCGAAGAAGCCCCGGTC 21597


RESULT 6
US-09-836-821-5
; Sequence 5, Application US/09836821
; Publication No. US20030087405A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/836,821
;  CURRENT FILING DATE:  2001-04-17
;  PRIOR APPLICATION NUMBER: 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 5
;   LENGTH: 36778
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-836-821-5

  Query Match           55.5%;  Score 715.4;  DB 9;  Length 36778;
  Best Local Similarity  72.6%;  Pred. No. 8.1e-165;
  Matches  939;  Conservative    0;  Mismatches  351;  Indels    3;  Gaps    1;

Qy          1 GAGCCGATCGCCATCGTTGGCATGGCCTGTCGTTTCCCCGGCGGAGTGACCTCGGCGGAC 60
              ||||| |||| ||||| |||||||| || ||  | || ||||| ||  ||||| ||||
Db      20305 GAGCCCGTCGCGATCGTCGGCATGGCGTGCCGCCTGCCGGGCGGTGTCGCCTCGCCGGAG 20364

Qy         61 GACTTCTGGGATCTGATCTCCTCCGAGCAGGACGCGATCGGCGGATTCCCCACCGACCGC 120
              ||| | |||   ||| |  ||  ||   |||||||||||  ||| |||||    ||||||
Db      20365 GACCTGTGGCGGCTGGTGGCCGGCGGCGAGGACGCGATCTCCGGCTTCCCGCAGGACCGC 20424

Qy        121 GGCTGGGACCTGGACACGCTCTACGACCCCGACCCCGACCACCCCGGCACCTGCTACACC 180
              |||||||| ||||   ||| |||||||| ||||| |||     |||          |||  |
Db      20425 GGCTGGGACGTGGAGGGGCTGTACGACCCGGACCCGGACGCGTCCGGGCGGACGTACTGC 20484

Qy        181 CGAAACGGCGGATTCCTCTACGACGCAGGCCACTTCGACGCCGAATTCTTCGGCATCAGC 240
              ||   ||| || |||||| |||| || ||| | ||||||||||| ||||||||  |||
Db      20485 CGTGCCGGTGGCTTCCTCGACGAGGCGGGCGAGTTCGACGCCGACTTCTTCGGGATCTCG 20544

Qy        241 CCCCGCGAAGCCCTCGCCATGGACCCCCAGCAACGACTCCTCCTCGAAACCGCCTGGGAA 300
              || ||||| |||||||||||||||||| ||||| || |||||||| || ||| |||||||
Db      20545 CCGCGCGAGGCCCTCGCCATGGACCCGCAGCAGCGGCTCCTCCTGGAGACCTCCTGGGAG 20604
```

```
Db      20545 CCGCGCGAGGCCCTCGCCATGGACCCGCAGCAGCGGCTCCTCCTGGAGACCTCCTGGGAG 20604

Qy        301 ACCATCGAACACGCCGGCATCAACCCCCACACCCTCCACGGCACCCCCACCGGAGTCTTC 360
              || ||||  ||||||| ||| |||||   | |||| || ||     |    ||| || |||
Db      20605 GCCGTCGAGGACGCCGGGATCGACCCGACCTCCCTTCAGGGGCAGCAGGTCGGCGTGTTC 20664

Qy        361 ACCGGCACCAACGGACAGGACTACGCACTTCGCGTGCACAACGCGGGCCAGTCAACCGAT 420
              | ||||||||||| |    ||||||  |  |    | | |||| | | | ||       ||
Db      20665 GCGGGCACCAACGGCCCCCACTACGAGCCGCTGCTCCGCAACACCGCCGAGGATCTTGAG 20724

Qy        421 GGTTTCGCACTGACCGGAACCGCCGGCAGCGTCATCTCCGGTCGTATCTCGTACACGTTT 480
              |||| ||    ||| || | ||||| |||| |||| || || ||| ||||||||||  |
Db      20725 GGTTACGTCGGGACGGGCAACGCCGCCAGCATCATGTCGGGCCGTGTCTCGTACACCCTC 20784

Qy        481 GGTTTTGAGGGTCCTGCGGTGTCGGTGGACACGGCTTGTTCCTCGTCGTTGGTGGCTTTG 540
              ||  | |||||  || || ||  |||| |||||  || || ||||| ||| |||| ||  ||
Db      20785 GGCCTGGAGGGCCCGGCCGTCACGGTCGACACCGCCTGCTCCTCCTCGCTGGTCGCCCTG 20844

Qy        541 CATCTGGCCTGTCAGGCGTTGCGTGCGGGTGAGTGCTCGATGGCGCTTGCCGGGGGTGTG 600
              || || |||   |||||  ||||   ||| || |||    ||||||| || || ||||||
Db      20845 CACCTCGCCGTGCAGGCCCTGCGCAAGGGCGAATGCGGACTGGCGCTCGCGGGCGGTGTG 20904

Qy        601 ACGGTGATGTCGTCTCCGGGTGCCTTCGTGGAGTTTTCGCGGCAGCGGGGTCTGGCCGCG 660
              ||||| ||||||  | ||      | ||||||||||    |||||||| || || || | |
Db      20905 ACGGTCATGTCGACGCCCACGACGTTCGTGGAGTTCAGCCGGCAGCGCGGGCTCGCGGAG 20964

Qy        661 GACGGGCATTGCAAGGCGTTCTCGGCGGCGGCGGACGGGACCGGCTGGGGTGAGGGTGTG 720
              ||||| |  |  |||||||||  | ||| ||||||||||   ||||  ||  ||||| ||
Db      20965 GACGGCCGGTCGAAGGCGTTCGCCGCGTCGGCGGACGGCTTCGGCCCGGCGGAGGGCGTC 21024

Qy        721 GGGATGCTGCTGGTGGAGCGGCTCTCCGACGCCCATCGCAACGGTCACCGTGTCCTGGCC 780
              || ||||| || || ||||| || || |||||||   |||||||| ||||||| |||||
Db      21025 GGCATGCTCCTCGTCGAGCGCCTGTCGGACGCCCGCCGCAACGGACACCGTGTGCTGGCG 21084

Qy        781 GTGGTGCGTGGCAGTGCGGTCAACCAGGACGGTGCGAGCAACGGTCTGACCGCGCCCAAC 840
              || ||||| |||||  |||||||||||||||||| |||||||||  |||||||| || |||
Db      21085 GTCGTGCGCGGCAGCGCGGTCAACCAGGACGGCGCGAGCAACGGCCTGACCGCCCCGAAC 21144

Qy        841 GGGCCGTCCCAGCAGCGTGTCATCCGCCAGGCCCTCGCCAACGCCGGCTTGTCGGCCGGT 900
              ||||| || |||||||| ||||||||  |  || |||||  |||||  |  || || |||
Db      21145 GGGCCCTCGCAGCAGCGCGTCATCCGGCGCGCGCTCGCGGACGCCCGACTGACGACCGCC 21204

Qy        901 GATGTCGACGCGGTGGAGGCCCACGGCACCGGCACCACTTTGGGCGACCCGATCGAGGCC 960
              || || |||| ||  |||||||||||||| |||||      | ||||||||||||||||||
Db      21205 GACGTGGACGTCGTCGAGGCCCACGGCACGGGCACGCGACTCGGCGACCCGATCGAGGCA 21264

Qy        961 CAGGCCCTCCTCGCGACCTACGGACAGGACCGTGCCGGCGAGGGGCCGCTGTGGCTGGGC 1020
              |||||||| |||| |||||||| ||||  || | |  |||   |||||| | |||||
Db      21265 CAGGCCCTCATCGCCACCTACGGCCAGGGGCGCGACACCGAACAGCCGCTGCGCCTGGGG 21324

Qy       1021 TCGGTCAAGTCCAATGTCGGTCACACACAGGCTGCCGCGGGCGTCGCCGGGGTGATCAAG 1080
              ||| | ||||||||  |||| |||||  ||||| ||||| || ||| ||||  | ||||||
Db      21325 TCGTTGAAGTCCAACATCGGACACACCCAGGCCGCCGCCGGTGTCTCCGGCATCATCAAG 21384

Qy       1081 ATGGTGATGGCGCTGCGGCATGGTCTGCTGCCGCGGACGTTGCATGTGGATGAGCCGTCG 1140
              |||||   |||| |||| || ||  | ||||||  |||| | || |||||   |||||||
Db      21385 ATGGTCCAGGCGATGCGCCACGGCGTCCTGCCGAAGACGCTCCACGTGGACCGGCCGTCG 21444

Qy       1141 CCGCATGTGGACTGGTCCGCGGGTGCGGTGCAGCTGCTGACGGAGACGGTGCCCTGGCCC 1200
                ||  | |||||||| |||||   ||||  ||||||| || ||| |  ||  ||||||
Db      21445 GACCAGATCGACTGGTCGGCGGGCACGGTCGAGCTGCTCACCGAGGCCATGGACTGGCCG 21504

Qy       1201 GG---CGGGGAGGGGCGGCTACGGCGGGCAGGAGTGTCATCATTCGGCGTCAGCGGCACC 1257
                |      ||||||  |||| || || || |  || || || |||||| ||||||||||
Db      21505 AGGAAGCAGGAGGGCGGGGCTGCGCCGCGCGGCCGTCTCCTCCTTCGGCATCAGCGGCACG 21564

Qy       1258 AACGCCCACGTCATCCTCGAAGAAGCACCCGCC 1290
              ||||| ||| || | |||||||||||| || | |
Db      21565 AACGCGCACATCGTGCTCGAAGAAGCCCCGGTC 21597

RESULT 7
US-09-861-289-5
```

```
; Sequence 5, Application US/09861289
; Patent No. US20020110897A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/861,289
;  CURRENT FILING DATE:  2001-05-18
;  PRIOR APPLICATION NUMBER: 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 5
;   LENGTH: 36778
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-861-289-5

  Query Match           55.5%;  Score 715.4;  DB 10;  Length 36778;
  Best Local Similarity   72.6%;  Pred. No. 8.1e-165;
  Matches  939;  Conservative    0;  Mismatches  351;  Indels    3;  Gaps    1;

Qy          1 GAGCCGATCGCCATCGTTGGCATGGCCTGTCGTTTCCCCGGCGGAGTGACCTCGGCGGAC 60
              ||||| |||| ||||| |||||||| || ||  | || ||||| ||  ||||| ||||
Db      20305 GAGCCCGTCGCGATCGTCGGCATGGCGTGCCGCCTGCCGGGCGGTGTCGCCTCGCCGGAG 20364

Qy         61 GACTTCTGGGATCTGATCTCCTCCGAGCAGGACGCGATCGGCGGATTCCCCACCGACCGC 120
              ||| | |||   ||| |  ||  ||  ||||||||||| ||| |||||    ||||||
Db      20365 GACCTGTGGCGGCTGGTGGCCGGCGGCGAGGACGCGATCTCCGGCTTCCCGCAGGACCGC 20424

Qy        121 GGCTGGGACCTGGACACGCTCTACGACCCCGACCCCGACCACCCCGGCACCTGCTACACC 180
              |||||||| ||||   ||| |||||||| ||||| |||   ||||         ||| |
Db      20425 GGCTGGGACGTGGAGGGGCTGTACGACCCGGACCCGGACGCGTCCGGGCGGACGTACTGC 20484

Qy        181 CGAAACGGCGGATTCCTCTACGACGCAGGCCACTTCGACGCCGAATTCTTCGGCATCAGC 240
              ||   ||| || |||||| |||| || ||| | |||||||||| ||||||| |||
Db      20485 CGTGCCGGTGGCTTCCTCGACGAGGCGGGCGAGTTCGACGCCGACTTCTTCGGGATCTCG 20544

Qy        241 CCCCGCGAAGCCCTCGCCATGGACCCCCAGCAACGACTCCTCCTCGAAACCGCCTGGGAA 300
              || ||||| |||||||||||||||||| ||||| || |||||||| || ||| |||||||
Db      20545 CCGCGCGAGGCCCTCGCCATGGACCCGCAGCAGCGGCTCCTCCTGGAGACCTCCTGGGAG 20604

Qy        301 ACCATCGAACACGCCGGCATCAACCCCCACACCCTCCACGGCACCCCCACCGGAGTCTTC 360
              || ||||  ||||||| ||| ||||    | ||||  || ||     |    ||| || |||
Db      20605 GCCGTCGAGGACGCCGGGATCGACCCGACCTCCCTTCAGGGGCAGCAGGTCGGCGTGTTC 20664

Qy        361 ACCGGCACCAACGGACAGGACTACGCACTTCGCGTGCACAACGCGGGCCAGTCAACCGAT 420
              | ||||||||||| |    ||||||  |  |    | | |||| | | | ||       ||
Db      20665 GCGGGCACCAACGGCCCCCACTACGAGCCGCTGCTCCGCAACACCGCCGAGGATCTTGAG 20724

Qy        421 GGTTTCGCACTGACCGGAACCGCCGGCAGCGTCATCTCCGGTCGTATCTCGTACACGTTT 480
              |||| ||    ||| || | ||||| |||| |||| || || ||| ||||||||||  |
Db      20725 GGTTACGTCGGGACGGGCAACGCCGCCAGCATCATGTCGGGCCGTGTCTCGTACACCCTC 20784

Qy        481 GGTTTTGAGGGTCCTGCGGTGTCGGTGGACACGGCTTGTTCCTCGTCGTTGGTGGCTTTG 540
              ||  | |||||  || || ||  |||| |||||  || || ||||| ||| |||| ||  ||
Db      20785 GGCCTGGAGGGCCCGGCCGTCACGGTCGACACCGCCTGCTCCTCCTCGCTGGTCGCCCTG 20844

Qy        541 CATCTGGCCTGTCAGGCGTTGCGTGCGGGTGAGTGCTCGATGGCGCTTGCCGGGGGTGTG 600
              || || |||   |||||  ||||   ||| || |||    ||||||| || || ||||||
Db      20845 CACCTCGCCGTGCAGGCCCTGCGCAAGGGCGAATGCGGACTGGCGCTCGCGGGCGGTGTG 20904

Qy        601 ACGGTGATGTCGTCTCCGGGTGCCTTCGTGGAGTTTTCGCGGCAGCGGGGTCTGGCCGCG 660
              ||||| ||||||  | ||      | ||||||||||    |||||||| || || || | |
Db      20905 ACGGTCATGTCGACGCCCACGACGTTCGTGGAGTTCAGCCGGCAGCGCGGGCTCGCGGAG 20964

Qy        661 GACGGGCATTGCAAGGCGTTCTCGGCGGCGGCGGACGGGACCGGCTGGGGTGAGGGTGTG 720
              ||||| |  |  |||||||||  | ||| ||||||||||   ||||  ||  ||||| ||
Db      20965 GACGGCCGGTCGAAGGCGTTCGCCGCGTCGGCGGACGGCTTCGGCCCGGCGGAGGGCGTC 21024
```

```
Qy        721 GGGATGCTGCTGGTGGAGCGGCTCTCCGACGCCCATCGCAACGGTCACCGTGTCCTGGCC 780
              || ||||| || || ||||| || || ||||||| |||||||| |||||||| |||||
Db      21025 GGCATGCTCCTCGTCGAGCGCCTGTCGGACGCCCGCCGCAACGGACACCGTGTGCTGGCG 21084

Qy        781 GTGGTGCGTGGCAGTGCGGTCAACCAGGACGGTGCGAGCAACGGTCTGACCGCGCCCAAC 840
              || ||||| ||||| |||||||||||||||||| |||||||||| |||||||| || |||
Db      21085 GTCGTGCGCGGCAGCGCGGTCAACCAGGACGGCGCGAGCAACGGCCTGACCGCCCCGAAC 21144

Qy        841 GGGCCGTCCCAGCAGCGTGTCATCCGCCAGGCCCTCGCCAACGCCGGCTTGTCGGCCGGT 900
              ||||| || |||||||| |||||||| |  || |||||  |||||  |  || ||| |||
Db      21145 GGGCCCTCGCAGCAGCGCGTCATCCGGCGCGCGCTCGCGGACGCCCGACTGACGACCGCC 21204

Qy        901 GATGTCGACGCGGTGGAGGCCCACGGCACCGGCACCACTTTGGGCGACCCGATCGAGGCC 960
              || || ||||  || ||||||||||||||| |||||     |||||||||||||||||||
Db      21205 GACGTGGACGTCGTCGAGGCCCACGGCACGGGCACGCGACTCGGCGACCCGATCGAGGCA 21264

Qy        961 CAGGCCCTCCTCGCGACCTACGGACAGGACCGTGCCGGCGAGGGGCCGCTGTGGCTGGGC 1020
              ||||||||| |||| |||||||| ||||  || | |  |||   ||||||| | |||||
Db      21265 CAGGCCCTCATCGCCACCTACGGCCAGGGGCGCGACACCGAACAGCCGCTGCGCCTGGGG 21324

Qy       1021 TCGGTCAAGTCCAATGTCGGTCACACACAGGCTGCCGCGGGCGTCGCCGGGGTGATCAAG 1080
              ||| | |||||||| |||| ||||| ||||| |||| || ||| |||| | ||||||
Db      21325 TCGTTGAAGTCCAACATCGGACACACCCAGGCCGCCGCCGGTGTCTCCGGCATCATCAAG 21384

Qy       1081 ATGGTGATGGCGCTGCGGCATGGTCTGCTGCCGCGGACGTTGCATGTGGATGAGCCGTCG 1140
              |||||   |||| |||| || ||  | |||||  |||| | || |||||   |||||||
Db      21385 ATGGTCCAGGCGATGCGCCACGGCGTCCTGCCGAAGACGCTCCACGTGGACCGGCCGTCG 21444

Qy       1141 CCGCATGTGGACTGGTCCGCGGGTGCGGTGCAGCTGCTGACGGAGACGGTGCCCTGGCCC 1200
                ||  | |||||||| |||||  ||||  ||||||| || ||| |  ||  ||||||
Db      21445 GACCAGATCGACTGGTCGGCGGGCACGGTCGAGCTGCTCACCGAGGCCATGGACTGGCCG 21504

Qy       1201 GG---CGGGGAGGGGCGGCTACGGCGGGCAGGAGTGTCATCATTCGGCGTCAGCGGCACC 1257
               |      ||||||  |||| || || || |  || || || |||||  ||||||||||
Db      21505 AGGAAGCAGGAGGGCGGGCTGCGCCGCGCGGCCGTCTCCTCCTTCGGCATCAGCGGCACG 21564

Qy       1258 AACGCCCACGTCATCCTCGAAGAAGCACCCGCC 1290
              ||||| ||| || | |||||||||||| || | |
Db      21565 AACGCGCACATCGTGCTCGAAGAAGCCCCGGTC 21597


RESULT 8
US-09-988-384B-5
; Sequence 5, Application US/09988384B
; Publication No. US20030073824A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.536US1
;  CURRENT APPLICATION NUMBER: US/09/988,384B
;  CURRENT FILING DATE:  2001-11-19
;  PRIOR APPLICATION NUMBER: PCT/US99/14398
;  PRIOR FILING DATE: 1999-06-25
;  PRIOR APPLICATION NUMBER: US 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 53
; SEQ ID NO 5
;   LENGTH: 37948
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-988-384B-5

  Query Match             55.5%;  Score 715.4;  DB 9;  Length 37948;
  Best Local Similarity   72.6%;  Pred. No. 8.1e-165;
  Matches  939;  Conservative    0;  Mismatches  351;  Indels    3;  Gaps    1;

Qy          1 GAGCCGATCGCCATCGTTGGCATGGCCTGTCGTTTCCCCGGCGGAGTGACCTCGGCGGAC 60
              |||||  |||| ||||| |||||||| || ||  | || ||||| ||  |||||  ||||
Db      21475 GAGCCCGTCGCGATCGTCGGCATGGCGTGCCGCCTGCCGGGCGGTGTCGCCTCGCCGGAG 21534

Qy         61 GACTTCTGGGATCTGATCTCCTCCGAGCAGGACGCGATCGGCGGATTCCCCACCGACCGC 120
              ||| | |||   ||| |  ||  ||   |||||||||| ||| |||||      ||||||
Db      21535 GACCTGTGGCGGCTGGTGGCCGGCGGCGAGGACGCGATCTCCGGCTTCCCGCAGGACCGC 21594

Qy        121 GGCTGGGACCTGGACACGCTCTACGACCCCGACCCCGACCACCCCGGCACCTGCTACACC 180
              |||||||| ||||   ||| |||||||| ||||| |||     ||||         ||| |
Db      21595 GGCTGGGACGTGGAGGGGCTGTACGACCCGGACCCGGACGCGTCCGGGCGGACGTACTGC 21654

Qy        181 CGAAACGGCGGATTCCTCTACGACGCAGGCCACTTCGACGCCGAATTCTTCGGCATCAGC 240
              ||   ||| || ||||| |||| || ||| | ||||||||||| ||||||| |||
Db      21655 CGTGCCGGTGGCTTCCTCGACGAGGCGGGCGAGTTCGACGCCGACTTCTTCGGGATCTCG 21714

Qy        241 CCCCGCGAAGCCCTCGCCATGGACCCCCAGCAACGACTCCTCCTCGAAACCGCCTGGGAA 300
              || ||||| |||||||||||||||| ||||| || ||||||| || ||| |||||||
Db      21715 CCGCGCGAGGCCCTCGCCATGGACCCGCAGCAGCGGCTCCTCCTGGAGACCTCCTGGGAG 21774

Qy        301 ACCATCGAACACGCCGGCATCAACCCCCACACCCTCCACGGCACCCCCACCGGAGTCTTC 360
              || ||||  ||||||| ||| ||||    | |||| || ||    |   ||| || |||
Db      21775 GCCGTCGAGGACGCCGGGATCGACCCGACCTCCCTTCAGGGGCAGCAGGTCGGCGTGTTC 21834

Qy        361 ACCGGCACCAACGGACAGGACTACGCACTTCGCGTGCACAACGCGGGCCAGTCAACCGAT 420
              | |||||||||| |  |||||  |  |  | | |||| | | ||  ||
Db      21835 GCGGGCACCAACGGCCCCCACTACGAGCCGCTGCTCCGCAACACCGCCGAGGATCTTGAG 21894

Qy        421 GGTTTCGCACTGACCGGAACCGCCGGCAGCGTCATCTCCGGTCGTATCTCGTACACGTTT 480
              |||| ||   ||| || | ||||| |||| |||| || || ||| ||||||||| |
Db      21895 GGTTACGTCGGGACGGGCAACGCCGCCAGCATCATGTCGGGCCGTGTCTCGTACACCCTC 21954

Qy        481 GGTTTTGAGGGTCCTGCGGTGTCGGTGGACACGGCTTGTTCCTCGTCGTTGGTGGCTTTG 540
              ||  | |||| || || ||  |||| |||| || || |||| ||| |||| ||  ||
Db      21955 GGCCTGGAGGGCCCGGCCGTCACGGTCGACACCGCCTGCTCCTCCTCGCTGGTCGCCCTG 22014

Qy        541 CATCTGGCCTGTCAGGCGTTGCGTGCGGGTGAGTGCTCGATGGCGCTTGCCGGGGGTGTG 600
              || || |||   |||||  |||  ||| || |||    ||||||| || || ||||||
Db      22015 CACCTCGCCGTGCAGGCCCTGCGCAAGGGCGAATGCGGACTGGCGCTCGCGGGCGGTGTG 22074

Qy        601 ACGGTGATGTCGTCTCCGGGTGCCTTCGTGGAGTTTTCGCGGCAGCGGGGTCTGGCCGCG 660
              ||||| |||||| | ||     | ||||||||||    ||||||| || || || | |
Db      22075 ACGGTCATGTCGACGCCCACGACGTTCGTGGAGTTCAGCCGGCAGCGCGGGCTCGCGGAG 22134

Qy        661 GACGGGCATTGCAAGGCGTTCTCGGCGGCGGCGGACGGGACCGGCTGGGGTGAGGGTGTG 720
              ||||| |  | ||||||||| | ||| ||||||||||    |||| ||  ||||| ||
Db      22135 GACGGCCGGTCGAAGGCGTTCGCCGCGTCGGCGGACGGCTTCGGCCCGGCGGAGGGCGTC 22194

Qy        721 GGGATGCTGCTGGTGGAGCGGCTCTCCGACGCCCATCGCAACGGTCACCGTGTCCTGGCC 780
              || ||||| || || ||||| || || ||||||| |||||||| |||||||| |||||
Db      22195 GGCATGCTCCTCGTCGAGCGCCTGTCGGACGCCCGCCGCAACGGACACCGTGTGCTGGCG 22254

Qy        781 GTGGTGCGTGGCAGTGCGGTCAACCAGGACGGTGCGAGCAACGGTCTGACCGCGCCCAAC 840
              || ||||| ||||| |||||||||||||||||| |||||||||| |||||||| || |||
Db      22255 GTCGTGCGCGGCAGCGCGGTCAACCAGGACGGCGCGAGCAACGGCCTGACCGCCCCGAAC 22314

Qy        841 GGGCCGTCCCAGCAGCGTGTCATCCGCCAGGCCCTCGCCAACGCCGGCTTGTCGGCCGGT 900
              ||||| || |||||||| |||||||| |  || |||||  |||||  |  || || |||
Db      22315 GGGCCCTCGCAGCAGCGCGTCATCCGGCGCGCGCTCGCGGACGCCCGACTGACGACCGCC 22374

Qy        901 GATGTCGACGCGGTGGAGGCCCACGGCACCGGCACCACTTTGGGCGACCCGATCGAGGCC 960
              || || ||||  || ||||||||||||||| |||||     |||||||||||||||||||
Db      22375 GACGTGGACGTCGTCGAGGCCCACGGCACGGGCACGCGACTCGGCGACCCGATCGAGGCA 22434

Qy        961 CAGGCCCTCCTCGCGACCTACGGACAGGACCGTGCCGGCGAGGGGCCGCTGTGGCTGGGC 1020
              ||||||||| |||| |||||||| |||  || | |  |||   ||||||| | |||||
Db      22435 CAGGCCCTCATCGCCACCTACGGCCAGGGGCGCGACACCGAACAGCCGCTGCGCCTGGGG 22494

Qy       1021 TCGGTCAAGTCCAATGTCGGTCACACACAGGCTGCCGCGGGCGTCGCCGGGGTGATCAAG 1080
              ||| | |||||||| |||| ||||| ||||| |||| || ||| |||| | ||||||
Db      22495 TCGTTGAAGTCCAACATCGGACACACCCAGGCCGCCGCCGGTGTCTCCGGCATCATCAAG 22554

Qy       1081 ATGGTGATGGCGCTGCGGCATGGTCTGCTGCCGCGGACGTTGCATGTGGATGAGCCGTCG 1140
              |||||   |||| |||| || ||  | ||||||  |||| | || |||||   |||||||
Db      22555 ATGGTCCAGGCGATGCGCCACGGCGTCCTGCCGAAGACGCTCCACGTGGACCGGCCGTCG 22614

Qy       1141 CCGCATGTGGACTGGTCCGCGGGTGCGGTGCAGCTGCTGACGGAGACGGTGCCCTGGCCC 1200
```

```
Db      22615 GACCAGATCGACTGGTCGGCGGGCACGGTCGAGCTGCTCACCGAGGCCATGGACTGGCCG 22674

Qy       1201 GG---CGGGGAGGGGCGGCTACGGCGGGCAGGAGTGTCATCATTCGGCGTCAGCGGCACC 1257

Db      22675 AGGAAGCAGGAGGGCGGGCTGCGCCGCGCGGCCGTCTCCTCCTTCGGCATCAGCGGCACG 22734

Qy       1258 AACGCCCACGTCATCCTCGAAGAAGCACCCGCC 1290

Db      22735 AACGCGCACATCGTGCTCGAAGAAGCCCCGGTC 22767


RESULT 9
US-09-793-708-19
; Sequence 19, Application US/09793708
; Publication No. US20030104597A1
; GENERAL INFORMATION:
;  APPLICANT: ASHLEY, Gary
;  APPLICANT:  BETLACH, Melanie C.
;  APPLICANT:  BETLACH, Mary C.
;  APPLICANT:  McDANIEL, Robert
;  APPLICANT:  TANG, Li
;  TITLE OF INVENTION: RECOMBINANT NARBONOLIDE POLYKETIDE SYNTHASE
;  FILE REFERENCE: 300622002121
;  CURRENT APPLICATION NUMBER: US/09/793,708
;  CURRENT FILING DATE:  2001-02-22
;  PRIOR APPLICATION NUMBER: US 09/657,440
;  PRIOR FILING DATE: 2000-09-07
;  PRIOR APPLICATION NUMBER: US 09/320,878
;  PRIOR FILING DATE: 1999-05-27
;  PRIOR APPLICATION NUMBER: US 09/141,908
;  PRIOR FILING DATE: 1998-08-28
;  PRIOR APPLICATION NUMBER: US 09/073,538
;  PRIOR FILING DATE: 1998-05-06
;  PRIOR APPLICATION NUMBER: US 08/846,247
;  PRIOR FILING DATE: 1997-04-30
;  PRIOR APPLICATION NUMBER: US 60/134,990
;  PRIOR FILING DATE: 1999-05-20
;  NUMBER OF SEQ ID NOS: 38
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 19
;   LENGTH: 38506
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-793-708-19

  Query Match             55.5%;  Score 715.4;  DB 9;  Length 38506;
  Best Local Similarity   72.6%;  Pred. No. 8.1e-165;
  Matches  939;  Conservative    0;  Mismatches  351;  Indels    3;  Gaps    1;

Qy          1 GAGCCGATCGCCATCGTTGGCATGGCCTGTCGTTTCCCCGGCGGAGTGACCTCGGCGGAC 60

Db      18447 GAGCCCGTCGCGATCGTCGGCATGGCGTGCCGCCTGCCGGGCGGTGTCGCCTCGCCGGAG 18506

Qy         61 GACTTCTGGGATCTGATCTCCTCCGAGCAGGACGCGATCGGCGGATTCCCCACCGACCGC 120

Db      18507 GACCTGTGGCGGCTGGTGGCCGGCGGCGAGGACGCGATCTCCGGCTTCCCGCAGGACCGC 18566

Qy        121 GGCTGGGACCTGGACACGCTCTACGACCCCGACCCCGACCACCCCGGCACCTGCTACACC 180

Db      18567 GGCTGGGACGTGGAGGGGCTGTACGACCCGGACCCGGACGCGTCCGGGCGGACGTACTGC 18626

Qy        181 CGAAACGGCGGATTCCTCTACGACGCAGGCCACTTCGACGCCGAATTCTTCGGCATCAGC 240

Db      18627 CGTGCCGGTGGCTTCCTCGACGAGGCGGGCGAGTTCGACGCCGACTTCTTCGGGATCTCG 18686

Qy        241 CCCCGCGAAGCCCTCGCCATGGACCCCCAGCAACGACTCCTCCTCGAAACCGCCTGGGAA 300

Db      18687 CCGCGCGAGGCCCTCGCCATGGACCCGCAGCAGCGGCTCCTCCTGGAGACCTCCTGGGAG 18746

Qy        301 ACCATCGAACACGCCGGCATCAACCCCCACACCCTCCACGGCACCCCCACCGGAGTCTTC 360

Db      18747 GCCGTCGAGGACGCCGGGATCGACCCGACCTCCCTTCAGGGGCAGCAGGTCGGCGTGTTC 18806

Qy        361 ACCGGCACCAACGGACAGGACTACGCACTTCGCGTGCACAACGCGGGCCAGTCAACCGAT 420

Db      18807 GCGGGCACCAACGGCCCCCACTACGAGCCGCTGCTCCGCAACACCGCCGAGGATCTTGAG 18866

Qy        421 GGTTTCGCACTGACCGGAACCGCCGGCAGCGTCATCTCCGGTCGTATCTCGTACACGTTT 480

Db      18867 GGTTACGTCGGGACGGGCAACGCCGCCAGCATCATGTCGGGCCGTGTCTCGTACACCCTC 18926

Qy        481 GGTTTTGAGGGTCCTGCGGTGTCGGTGGACACGGCTTGTTCCTCGTCGTTGGTGGCTTTG 540

Db      18927 GGCCTGGAGGGCCCGGCCGTCACGGTCGACACCGCCTGCTCCTCCTCGCTGGTCGCCCTG 18986

Qy        541 CATCTGGCCTGTCAGGCGTTGCGTGCGGGTGAGTGCTCGATGGCGCTTGCCGGGGGTGTG 600

Db      18987 CACCTCGCCGTGCAGGCCCTGCGCAAGGGCGAATGCGGACTGGCGCTCGCGGGCGGTGTG 19046

Qy        601 ACGGTGATGTCGTCTCCGGGTGCCTTCGTGGAGTTTTCGCGGCAGCGGGGTCTGGCCGCG 660

Db      19047 ACGGTCATGTCGACGCCCACGACGTTCGTGGAGTTCAGCCGGCAGCGCGGGCTCGCGGAG 19106

Qy        661 GACGGGCATTGCAAGGCGTTCTCGGCGGCGGCGGACGGGACCGGCTGGGGTGAGGGTGTG 720

Db      19107 GACGGCCGGTCGAAGGCGTTCGCCGCGTCGGCGGACGGCTTCGGCCCGGCGGAGGGCGTC 19166

Qy        721 GGGATGCTGCTGGTGGAGCGGCTCTCCGACGCCCATCGCAACGGTCACCGTGTCCTGGCC 780

Db      19167 GGCATGCTCCTCGTCGAGCGCCTGTCGGACGCCCGCCGCAACGGACACCGTGTGCTGGCG 19226

Qy        781 GTGGTGCGTGGCAGTGCGGTCAACCAGGACGGTGCGAGCAACGGTCTGACCGCGCCCAAC 840

Db      19227 GTCGTGCGCGGCAGCGCGGTCAACCAGGACGGCGCGAGCAACGGCCTGACCGCCCCGAAC 19286

Qy        841 GGGCCGTCCCAGCAGCGTGTCATCCGCCAGGCCCTCGCCAACGCCGGCTTGTCGGCCGGT 900

Db      19287 GGGCCCTCGCAGCAGCGCGTCATCCGGCGCGCGCTCGCGGACGCCCGACTGACGACCGCC 19346

Qy        901 GATGTCGACGCGGTGGAGGCCCACGGCACCGGCACCACTTTGGGCGACCCGATCGAGGCC 960

Db      19347 GACGTGGACGTCGTCGAGGCCCACGGCACGGGCACGCGACTCGGCGACCCGATCGAGGCA 19406

Qy        961 CAGGCCCTCCTCGCGACCTACGGACAGGACCGTGCCGGCGAGGGGCCGCTGTGGCTGGGC 1020

Db      19407 CAGGCCCTCATCGCCACCTACGGCCAGGGGCGCGACACCGAACAGCCGCTGCGCCTGGGG 19466

Qy       1021 TCGGTCAAGTCCAATGTCGGTCACACACAGGCTGCCGCGGGCGTCGCCGGGGTGATCAAG 1080

Db      19467 TCGTTGAAGTCCAACATCGGACACACCCAGGCCGCCGCCGGTGTCTCCGGCATCATCAAG 19526

Qy       1081 ATGGTGATGGCGCTGCGGCATGGTCTGCTGCCGCGGACGTTGCATGTGGATGAGCCGTCG 1140

Db      19527 ATGGTCCAGGCGATGCGCCACGGCGTCCTGCCGAAGACGCTCCACGTGGACCGGCCGTCG 19586

Qy       1141 CCGCATGTGGACTGGTCCGCGGGTGCGGTGCAGCTGCTGACGGAGACGGTGCCCTGGCCC 1200

Db      19587 GACCAGATCGACTGGTCGGCGGGCACGGTCGAGCTGCTCACCGAGGCCATGGACTGGCCG 19646

Qy       1201 GG---CGGGGAGGGGCGGCTACGGCGGGCAGGAGTGTCATCATTCGGCGTCAGCGGCACC 1257

Db      19647 AGGAAGCAGGAGGGCGGGCTGCGCCGCGCGGCCGTCTCCTCCTTCGGCATCAGCGGCACG 19706

Qy       1258 AACGCCCACGTCATCCTCGAAGAAGCACCCGCC 1290

Db      19707 AACGCGCACATCGTGCTCGAAGAAGCCCCGGTC 19739


RESULT 10
US-09-860-846-30
; Sequence 30, Application US/09860846
; Patent No. US20020164742A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
```

```
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/860,846
;  CURRENT FILING DATE:  2001-05-18
;  PRIOR APPLICATION NUMBER: 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 30
;   LENGTH: 13842
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-860-846-30

  Query Match             53.4%;  Score 689.4;  DB 9;  Length 13842;
  Best Local Similarity   71.4%;  Pred. No. 1.7e-158;
  Matches  922;  Conservative    0;  Mismatches  366;  Indels    3;  Gaps    1;

Qy          1 GAGCCGATCGCCATCGTTGGCATGGCCTGTCGTTTCCCCGGCGGAGTGACCTCGGCGGAC 60
              |||||| | || ||||| |||||||||||| ||  | ||||| || ||  ||||| ||||
Db       3325 GAGCCGGTGGCGATCGTCGGCATGGCCTGCCGCCTGCCCGGTGGGGTCGCCTCGCCGGAG 3384

Qy         61 GACTTCTGGGATCTGATCTCCTCCGAGCAGGACGCGATCGGCGGATTCCCCACCGACCGC 120
              ||| | |||   ||| |  ||  ||   ||||||||||  |  |||||   ||||||
Db       3385 GACCTGTGGCGGCTGGTGGCCGGCGGCGGGGACGCGATCTCGGAGTTCCCGCAGGACCGC 3444

Qy        121 GGCTGGGACCTGGACACGCTCTACGACCCCGACCCCGACCACCCCGGCACCTGCTACACC 180
              ||||||||| ||||   ||| ||| |||| || || || |||||||||| |  |||  |
Db       3445 GGCTGGGACGTGGAGGGGCTGTACCACCCGGATCCGGAGCACCCCGGCACGTCGTACGTC 3504

Qy        181 CGAAACGGCGGATTCCTCTACGACGCAGGCCACTTCGACGCCGAATTCTTCGGCATCAGC 240
              ||  | ||||| ||| || |  |||  | |  |||||||| |  |||||||| |||
Db       3505 CGCCAGGGCGGTTTCATCGAGAACGTCGCCGGCTTCGACGCGGCCTTCTTCGGGATCTCG 3564

Qy        241 CCCCGCGAAGCCCTCGCCATGGACCCCCAGCAACGACTCCTCCTCGAAACCGCCTGGGAA 300
              || ||||| |||||||||||||||||| ||||| || ||||||||||||||| |||||||
Db       3565 CCGCGCGAGGCCCTCGCCATGGACCCGCAGCAGCGGCTCCTCCTCGAAACCTCCTGGGAG 3624

Qy        301 ACCATCGAACACGCCGGCATCAACCCCCACACCCTCCACGGCACCCCCACCGGAGTCTTC 360
              || ||||  ||||||| ||| ||||   | |||| |  ||     |   ||| ||||||
Db       3625 GCCGTCGAGGACGCCGGGATCGACCCGACCTCCCTGCGGGGACGGCAGGTCGGCGTCTTC 3684

Qy        361 ACCGGCACCAACGGACAGGACTACGCACTTCGCGTGCACAACGCGGGCCAGTCAACCGAT 420
              || ||  | |      || || |||| |   || |||   |||  ||  |      |||
Db       3685 ACTGGGGCGATGACCCACGAGTACGGGCCGAGCCTGCGGGACGGCGGGGAAGGCCTCGAC 3744

Qy        421 GGTTTCGCACTGACCGGAACCGCCGGCAGCGTCATCTCCGGTCGTATCTCGTACACGTTT 480
              ||  |  |||||||| | | | | |||||| || || || ||  ||||||||||  |
Db       3745 GGCTACCTGCTGACCGGCAACACGGCCAGCGTGATGTCGGGCCGCGTCTCGTACACACTC 3804

Qy        481 GGTTTTGAGGGTCCTGCGGTGTCGGTGGACACGGCTTGTTCCTCGTCGTTGGTGGCTTTG 540
              ||  ||||||| || ||  || |||||||||||||| || || |||||| |||| ||  ||
Db       3805 GGCCTTGAGGGCCCCGCCCTGACGGTGGACACGGCCTGCTCGTCGTCGCTGGTCGCCCTG 3864

Qy        541 CATCTGGCCTGTCAGGCGTTGCGTGCGGGTGAGTGCTCGATGGCGCTTGCCGGGGGTGTG 600
              || || |||   |||||  ||||   ||| |||  |   |||||||| ||||| || |||
Db       3865 CACCTCGCCGTGCAGGCCCTGCGCAAGGGCGAGGTCGACATGGCGCTCGCCGGCGGCGTG 3924

Qy        601 ACGGTGATGTCGTCTCCGGGTGCCTTCGTGGAGTTTTCGCGGCAGCGGGGTCTGGCCGCG 660
              | ||||||  |  | || ||    ||||| |||||    |||||||| || ||||||| |
Db       3925 GCCGTGATGCCCACGCCCGGGATGTTCGTCGAGTTCAGCCGGCAGCGCGGGCTGGCCGGG 3984

Qy        661 GACGGGCATTGCAAGGCGTTCTCGGCGGCGGCGGACGGGACCGGCTGGGGTGAGGGTGTG 720
              |||||  |  | ||||||||| | ||| |||||||||| ||| |||||    ||||| ||
Db       3985 GACGGCCGGTCGAAGGCGTTCGCCGCGTCGGCGGACGGCACCAGCTGGTCCGAGGGCGTC 4044

Qy        721 GGGATGCTGCTGGTGGAGCGGCTCTCCGACGCCCATCGCAACGGTCACCGTGTCCTGGCC 780
              ||  | || || || ||||| || || |||||||  |||||||| ||||  ||||| ||
Db       4045 GGCGTCCTCCTCGTCGAGCGCCTGTCGGACGCCCGCCGCAACGGACACCAGGTCCTCGCG 4104

Qy        781 GTGGTGCGTGGCAGTGCGGTCAACCAGGACGGTGCGAGCAACGGTCTGACCGCGCCCAAC 840
              || || || ||||| ||  | |||||||||| |||||||||| || || || || |||
Db       4105 GTCGTCCGCGGCAGCGCCTTGAACCAGGACGGCGCGAGCAACGGCCTCACGGCTCCGAAC 4164

Qy        841 GGGCCGTCCCAGCAGCGTGTCATCCGCCAGGCCCTCGCCAACGCCGGCTTGTCGGCCGGT 900
              ||||| || |||||||| |||||||| |  || || ||  |||||  |  || || ||
Db       4165 GGGCCCTCGCAGCAGCGCGTCATCCGGCGCGCGCTGGCGGACGCCCGGCTGACGACCTCC 4224

Qy        901 GATGTCGACGCGGTGGAGGCCCACGGCACCGGCACCACTTTGGGCGACCCGATCGAGGCC 960
              || || ||||  || ||||| |||||||| |||||      | ||||||||||||||||
Db       4225 GACGTGGACGTCGTCGAGGCACACGGCACGGGCACGCGACTCGGCGACCCGATCGAGGCG 4284

Qy        961 CAGGCCCTCCTCGCGACCTACGGACAGGACCGTGCCGGCGAGGGGCCGCTGTGGCTGGGC 1020
              ||||||||  |||| |||||||| |||| ||||| || |||   |||||||  | || ||
Db       4285 CAGGCCCTGATCGCCACCTACGGCCAGGGCCGTGACGACGAACAGCCGCTGCGCCTCGGG 4344

Qy       1021 TCGGTCAAGTCCAATGTCGGTCACACACAGGCTGCCGCGGGCGTCGCCGGGGTGATCAAG 1080
              ||| | ||||||||  |||| |||| ||||| || || |||||| |||| || ||||||
Db       4345 TCGTTGAAGTCCAACATCGGGCACACCCAGGCCGCGGCCGGCGTCTCCGGTGTCATCAAG 4404

Qy       1081 ATGGTGATGGCGCTGCGGCATGGTCTGCTGCCGCGGACGTTGCATGTGGATGAGCCGTCG 1140
              |||||  |||| |||| || || |||||||||  |||| |||| || || ||||| |||
Db       4405 ATGGTCCAGGCGATGCGCCACGGACTGCTGCCGAAGACGCTGCACGTCGACGAGCCCTCG 4464

Qy       1141 CCGCATGTGGACTGGTCCGCGGGTGCGGTGCAGCTGCTGACGGAGACGGTGCCCTGGCCC 1200
                ||  | |||||||| || || || || || || || |||  | ||  ||||||
Db       4465 GACCAGATCGACTGGTCGGCTGGCGCCGTGGAACTCCTCACCGAGGCCGTCGACTGGCCG 4524

Qy       1201 G--*-GCGGGGAGGGGCGGCTACGGCGGGCAGGAGTGTCATCATTCGGCGTCAGCGGCACC 1257
              |   ||| ||  |||| || ||||| |  || || || |||||  ||||||||||
Db       4525 GAGAAGCAGGACGGCGGGCTGCGCCGGGCCGCCGTCTCCTCCTTCGGGATCAGCGGCACC 4584

Qy       1258 AACGCCCACGTCATCCTCGAAGAAGCACCCG 1288
              || || || ||  | |||||||| || || |
Db       4585 AATGCGCATGTGGTGCTCGAAGAGGCCCCGG 4615


RESULT 11
US-09-988-384B-30
; Sequence 30, Application US/09988384B
; Publication No. US20030073824A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.536US1
;  CURRENT APPLICATION NUMBER: US/09/988,384B
;  CURRENT FILING DATE:  2001-11-19
;  PRIOR APPLICATION NUMBER: PCT/US99/14398
;  PRIOR FILING DATE: 1999-06-25
;  PRIOR APPLICATION NUMBER: US 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 53
; SEQ ID NO 30
;   LENGTH: 13842
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-988-384B-30

  Query Match             53.4%;  Score 689.4;  DB 9;  Length 13842;
  Best Local Similarity   71.4%;  Pred. No. 1.7e-158;
  Matches  922;  Conservative    0;  Mismatches  366;  Indels    3;  Gaps    1;

Qy          1 GAGCCGATCGCCATCGTTGGCATGGCCTGTCGTTTCCCCGGCGGAGTGACCTCGGCGGAC 60
              |||||| | || ||||| |||||||||||| ||  | ||||| || ||  ||||| ||||
Db       3325 GAGCCGGTGGCGATCGTCGGCATGGCCTGCCGCCTGCCCGGTGGGGTCGCCTCGCCGGAG 3384

Qy         61 GACTTCTGGGATCTGATCTCCTCCGAGCAGGACGCGATCGGCGGATTCCCCACCGACCGC 120
              ||| | |||   ||| |  ||  ||   ||||||||||  |  |||||   ||||||
Db       3385 GACCTGTGGCGGCTGGTGGCCGGCGGCGGGGACGCGATCTCGGAGTTCCCGCAGGACCGC 3444

Qy        121 GGCTGGGACCTGGACACGCTCTACGACCCCGACCCCGACCACCCCGGCACCTGCTACACC 180
```

```
Db       3445 GGCTGGGACGTGGAGGGGCTGTACCACCCGGATCCGGAGCACCCCGGCACGTCGTACGTC 3504

Qy        181 CGAAACGGCGGATTCCTCTACGACGCAGGCCACTTCGACGCCGAATTCTTCGGCATCAGC 240

Db       3505 CGCCAGGGCGGTTTCATCGAGAACGTCGCCGGCTTCGACGCGGCCTTCTTCGGGATCTCG 3564

Qy        241 CCCCGCGAAGCCCTCGCCATGGACCCCCAGCAACGACTCCTCCTCGAAACCGCCTGGGAA 300

Db       3565 CCGCGCGAGGCCCTCGCCATGGACCCGCAGCAGCGGCTCCTCCTCGAAACCTCCTGGGAG 3624

Qy        301 ACCATCGAACACGCCGGCATCAACCCCCACACCCTCCACGGCACCCCCACCGGAGTCTTC 360

Db       3625 GCCGTCGAGGACGCCGGGATCGACCCGACCTCCCTGCGGGGACGGCAGGTCGGCGTCTTC 3684

Qy        361 ACCGGCACCAACGGACAGGACTACGCACTTCGCGTGCACAACGCGGGCCAGTCAACCGAT 420

Db       3685 ACTGGGGCGATGACCCACGAGTACGGGCCGAGCCTGCGGGACGGCGGGGAAGGCCTCGAC 3744

Qy        421 GGTTTCGCACTGACCGGAACCGCCGGCAGCGTCATCTCCGGTCGTATCTCGTACACGTTT 480

Db       3745 GGCTACCTGCTGACCGGCAACACGGCCAGCGTGATGTCGGGCCGCGTCTCGTACACACTC 3804

Qy        481 GGTTTTGAGGGTCCTGCGGTGTCGGTGGACACGGCTTGTTCCTCGTCGTTGGTGGCTTTG 540

Db       3805 GGCCTTGAGGGCCCCGCCCTGACGGTGGACACGGCCTGCTCGTCGTCGCTGGTCGCCCTG 3864

Qy        541 CATCTGGCCTGTCAGGCGTTGCGTGCGGGTGAGTGCTCGATGGCGCTTGCCGGGGGTGTG 600

Db       3865 CACCTCGCCGTGCAGGCCCTGCGCAAGGGCGAGGTCGACATGGCGCTCGCCGGCGGCGTG 3924

Qy        601 ACGGTGATGTCGTCTCCGGGTGCCTTCGTGGAGTTTTCGCGGCAGCGGGGTCTGGCCGCG 660

Db       3925 GCCGTGATGCCCACGCCCGGGATGTTCGTCGAGTTCAGCCGGCAGCGCGGGCTGGCCGGG 3984

Qy        661 GACGGGCATTGCAAGGCGTTCTCGGCGGCGGCGGACGGGACCGGCTGGGGTGAGGGTGTG 720

Db       3985 GACGGCCGGTCGAAGGCGTTCGCCGCGTCGGCGGACGGCACCAGCTGGTCCGAGGGCGTC 4044

Qy        721 GGGATGCTGCTGGTGGAGCGGCTCTCCGACGCCCATCGCAACGGTCACCGTGTCCTGGCC 780

Db       4045 GGCGTCCTCCTCGTCGAGCGCCTGTCGGACGCCCGCCGCAACGGACACCAGGTCCTCGCG 4104

Qy        781 GTGGTGCGTGGCAGTGCGGTCAACCAGGACGGTGCGAGCAACGGTCTGACCGCGCCCAAC 840

Db       4105 GTCGTCCGCGGCAGCGCCTTGAACCAGGACGGCGCGAGCAACGGCCTCACGGCTCCGAAC 4164

Qy        841 GGGCCGTCCCAGCAGCGTGTCATCCGCCAGGCCCTCGCCAACGCCGGCTTGTCGGCCGGT 900

Db       4165 GGGCCCTCGCAGCAGCGCGTCATCCGGCGCGCGCTGGCGGACGCCCGGCTGACGACCTCC 4224

Qy        901 GATGTCGACGCGGTGGAGGCCCACGGCACCGGCACCACTTTGGGCGACCCGATCGAGGCC 960

Db       4225 GACGTGGACGTCGTCGAGGCACACGGCACGGGCACGCGACTCGGCGACCCGATCGAGGCG 4284

Qy        961 CAGGCCCTCCTCGCGACCTACGGACAGGACCGTGCCGGCGAGGGGCCGCTGTGGCTGGGC 1020

Db       4285 CAGGCCCTGATCGCCACCTACGGCCAGGGCCGTGACGACGAACAGCCGCTGCGCCTCGGG 4344

Qy       1021 TCGGTCAAGTCCAATGTCGGTCACACACAGGCTGCCGCGGGCGTCGCCGGGGTGATCAAG 1080

Db       4345 TCGTTGAAGTCCAACATCGGGCACACCCAGGCCGCGGCCGGCGTCTCCGGTGTCATCAAG 4404

Qy       1081 ATGGTGATGGCGCTGCGGCATGGTCTGCTGCCGCGGACGTTGCATGTGGATGAGCCGTCG 1140

Db       4405 ATGGTCCAGGCGATGCGCCACGGACTGCTGCCGAAGACGCTGCACGTCGACGAGCCCTCG 4464

Qy       1141 CCGCATGTGGACTGGTCCGCGGGTGCGGTGCAGCTGCTGACGGAGACGGTGCCCTGGCCC 1200

Db       4465 GACCAGATCGACTGGTCGGCTGGCGCCGTGGAACTCCTCACCGAGGCCGTCGACTGGCCG 4524

Qy       1201 G---GCGGGGAGGGGCGGCTACGGCGGGCAGGAGTGTCATCATTCGGCGTCAGCGGCACC 1257

Db       4525 GAGAAGCAGGACGGCGGGCTGCGCCGGGCCGCCGTCTCCTCCTTCGGGATCAGCGGCACC 4584

Qy       1258 AACGCCCACGTCATCCTCGAAGAAGCACCCG 1288

Db       4585 AATGCGCATGTGGTGCTCGAAGAGGCCCCGG 4615


RESULT 12
US-09-836-821-30
; Sequence 30, Application US/09836821
; Publication No. US20030087405A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/836,821
;  CURRENT FILING DATE:  2001-04-17
;  PRIOR APPLICATION NUMBER: 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 30
;   LENGTH: 13842
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-836-821-30

  Query Match           53.4%;  Score 689.4;  DB 9;  Length 13842;
  Best Local Similarity  71.4%;  Pred. No. 1.7e-158;
  Matches  922;  Conservative    0;  Mismatches  366;  Indels    3;  Gaps    1;

Qy          1 GAGCCGATCGCCATCGTTGGCATGGCCTGTCGTTTCCCCGGCGGAGTGACCTCGGCGGAC 60

Db       3325 GAGCCGGTGGCGATCGTCGGCATGGCCTGCCGCCTGCCCGGTGGGGTCGCCTCGCCGGAG 3384

Qy         61 GACTTCTGGGATCTGATCTCCTCCGAGCAGGACGCGATCGGCGGATTCCCCACCGACCGC 120

Db       3385 GACCTGTGGCGGCTGGTGGCCGGCGGCGGGGACGCGATCTCGGAGTTCCCGCAGGACCGC 3444

Qy        121 GGCTGGGACCTGGACACGCTCTACGACCCCGACCCCGACCACCCCGGCACCTGCTACACC 180

Db       3445 GGCTGGGACGTGGAGGGGCTGTACCACCCGGATCCGGAGCACCCCGGCACGTCGTACGTC 3504

Qy        181 CGAAACGGCGGATTCCTCTACGACGCAGGCCACTTCGACGCCGAATTCTTCGGCATCAGC 240

Db       3505 CGCCAGGGCGGTTTCATCGAGAACGTCGCCGGCTTCGACGCGGCCTTCTTCGGGATCTCG 3564

Qy        241 CCCCGCGAAGCCCTCGCCATGGACCCCCAGCAACGACTCCTCCTCGAAACCGCCTGGGAA 300

Db       3565 CCGCGCGAGGCCCTCGCCATGGACCCGCAGCAGCGGCTCCTCCTCGAAACCTCCTGGGAG 3624

Qy        301 ACCATCGAACACGCCGGCATCAACCCCCACACCCTCCACGGCACCCCCACCGGAGTCTTC 360

Db       3625 GCCGTCGAGGACGCCGGGATCGACCCGACCTCCCTGCGGGGACGGCAGGTCGGCGTCTTC 3684

Qy        361 ACCGGCACCAACGGACAGGACTACGCACTTCGCGTGCACAACGCGGGCCAGTCAACCGAT 420

Db       3685 ACTGGGGCGATGACCCACGAGTACGGGCCGAGCCTGCGGGACGGCGGGGAAGGCCTCGAC 3744

Qy        421 GGTTTCGCACTGACCGGAACCGCCGGCAGCGTCATCTCCGGTCGTATCTCGTACACGTTT 480

Db       3745 GGCTACCTGCTGACCGGCAACACGGCCAGCGTGATGTCGGGCCGCGTCTCGTACACACTC 3804

Qy        481 GGTTTTGAGGGTCCTGCGGTGTCGGTGGACACGGCTTGTTCCTCGTCGTTGGTGGCTTTG 540

Db       3805 GGCCTTGAGGGCCCCGCCCTGACGGTGGACACGGCCTGCTCGTCGTCGCTGGTCGCCCTG 3864

Qy        541 CATCTGGCCTGTCAGGCGTTGCGTGCGGGTGAGTGCTCGATGGCGCTTGCCGGGGGTGTG 600

Db       3865 CACCTCGCCGTGCAGGCCCTGCGCAAGGGCGAGGTCGACATGGCGCTCGCCGGCGGCGTG 3924
```

```
Qy        601 ACGGTGATGTCGTCTCCGGGTGCCTTCGTGGAGTTTTCGCGGCAGCGGGGTCTGGCCGCG 660
              | |||||| |   | || ||     ||||| |||||     |||||||| || ||||||| |
Db       3925 GCCGTGATGCCCACGCCCGGGATGTTCGTCGAGTTCAGCCGGCAGCGCGGGCTGGCCGGG 3984

Qy        661 GACGGGCATTGCAAGGCGTTCTCGGCGGCGGCGGACGGGACCGGCTGGGGTGAGGGTGTG 720
              |||||  |   | |||||||| | ||| |||||||||| ||| |||||    ||||| ||
Db       3985 GACGGCCGGTCGAAGGCGTTCGCCGCGTCGGCGGACGGCACCAGCTGGTCCGAGGGCGTC 4044

Qy        721 GGGATGCTGCTGGTGGAGCGGCTCTCCGACGCCCATCGCAACGGTCACCGTGTCCTGGCC 780
              ||  | || || || ||||| || || |||||||  ||||||| ||||  |||||  ||
Db       4045 GGCGTCCTCCTCGTCGAGCGCCTGTCGGACGCCCGCCGCAACGGACACCAGGTCCTCGCG 4104

Qy        781 GTGGTGCGTGGCAGTGCGGTCAACCAGGACGGTGCGAGCAACGGTCTGACCGCGCCCAAC 840
              || || || ||||| ||  | |||||||||| |||||||||| || || || || |||
Db       4105 GTCGTCCGCGGCAGCGCCTTGAACCAGGACGGCGCGAGCAACGGCCTCACGGCTCCGAAC 4164

Qy        841 GGGCCGTCCCAGCAGCGTGTCATCCGCCAGGCCCTCGCCAACGCCGGCTTGTCGGCCGGT 900
              ||||| || |||||||| |||||||| |  || || ||  |||||  |  || || ||
Db       4165 GGGCCCTCGCAGCAGCGCGTCATCCGGCGCGCGCTGGCGGACGCCCGGCTGACGACCTCC 4224

Qy        901 GATGTCGACGCGGTGGAGGCCCACGGCACCGGCACCACTTTGGGCGACCCGATCGAGGCC 960
              || || ||||  || ||||| |||||||| |||||      | ||||||||||||||||||
Db       4225 GACGTGGACGTCGTCGAGGCACACGGCACGGGCACGCGACTCGGCGACCCGATCGAGGCG 4284

Qy        961 CAGGCCCTCCTCGCGACCTACGGACAGGACCGTGCCGGCGAGGGGCCGCTGTGGCTGGGC 1020
              ||||||||  |||| |||||||| |||| ||||| || |||   |||||||  | || ||
Db       4285 CAGGCCCTGATCGCCACCTACGGCCAGGGCCGTGACGACGAACAGCCGCTGCGCCTCGGG 4344

Qy       1021 TCGGTCAAGTCCAATGTCGGTCACACACAGGCTGCCGCGGGCGTCGCCGGGGTGATCAAG 1080
              ||| | ||||||||  |||| ||||| ||||| || || ||||||  |||| || ||||||
Db       4345 TCGTTGAAGTCCAACATCGGGCACACCCAGGCCGCGGCCGGCGTCTCCGGTGTCATCAAG 4404

Qy       1081 ATGGTGATGGCGCTGCGGCATGGTCTGCTGCCGCGGACGTTGCATGTGGATGAGCCGTCG 1140
              |||||   |||| |||| || || |||||||||  |||| |||| || || ||||| |||
Db       4405 ATGGTCCAGGCGATGCGCCACGGACTGCTGCCGAAGACGCTGCACGTCGACGAGCCCTCG 4464

Qy       1141 CCGCATGTGGACTGGTCCGCGGGTGCGGTGCAGCTGCTGACGGAGACGGTGCCCTGGCCC 1200
                 |  | |||||||| || || || ||| | || || || ||| | ||   ||||||
Db       4465 GACCAGATCGACTGGTCGGCTGGCGCCGTGGAACTCCTCACCGAGGCCGTCGACTGGCCG 4524

Qy       1201 G---GCGGGGAGGGGCGGCTACGGCGGGCAGGAGTGTCATCATTCGGCGTCAGCGGCACC 1257
              |       ||| ||  |||| || ||||| |  || || || |||||  ||||||||||
Db       4525 GAGAAGCAGGACGGCGGGCTGCGCCGGGCCGCCGTCTCCTCCTTCGGGATCAGCGGCACC 4584

Qy       1258 AACGCCCACGTCATCCTCGAAGAAGCACCCG 1288
              || || || ||  | |||||||| || || |
Db       4585 AATGCGCATGTGGTGCTCGAAGAGGCCCCGG 4615
```

RESULT 13
US-09-861-289-30
; Sequence 30, Application US/09861289
; Patent No. US20020110897A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/861,289
; CURRENT FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 30
; LENGTH: 13842
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-861-289-30

Query Match 53.4%; Score 689.4; DB 10; Length 13842;
Best Local Similarity 71.4%; Pred. No. 1.7e-158;
Matches 922; Conservative 0; Mismatches 366; Indels 3; Gaps 1;

```
Qy          1 GAGCCGATCGCCATCGTTGGCATGGCCTGTCGTTTCCCCGGCGGAGTGACCTCGGCGGAC 60
              |||||| | || ||||| ||||||||||| ||  | ||||| || ||  ||||| ||||
Db       3325 GAGCCGGTGGCGATCGTCGGCATGGCCTGCCGCCTGCCCGGTGGGGTCGCCTCGCCGGAG 3384

Qy         61 GACTTCTGGGATCTGATCTCCTCCGAGCAGGACGCGATCGGCGGATTCCCCACCGACCGC 120
              ||| | |||    ||| |  ||  ||     ||||||||||    |  |||||     ||||||
Db       3385 GACCTGTGGCGGCTGGTGGCCGGCGGCGGGGACGCGATCTCGGAGTTCCCGCAGGACCGC 3444

Qy        121 GGCTGGGACCTGGACACGCTCTACGACCCCGACCCCGACCACCCCCGGCACCTGCTACACC 180
              |||||||||  ||||    ||| |||  |||| || || || ||||||||||  |   |||  |
Db       3445 GGCTGGGACGTGGAGGGGCTGTACCACCCGGATCCGGAGCACCCCGGCACGTCGTACGTC 3504

Qy        181 CGAAACGGCGGATTCCTCTACGACGCAGGCCACTTCGACGCCGAATTCTTCGGCATCAGC 240
              ||  | |||||  ||| ||  |  |||  | |  |||||||||  |  |||||||  |||
Db       3505 CGCCAGGGCGGTTTCATCGAGAACGTCGCCGGCTTCGACGCGGCCTTCTTCGGGATCTCG 3564

Qy        241 CCCCGCGAAGCCCTCGCCATGGACCCCCAGCAACGACTCCTCCTCGAAACCGCCTGGGAA 300
              || |||||  |||||||||||||||||  |||||  ||  ||||||||||||||  |||||||
Db       3565 CCGCGCGAGGCCCTCGCCATGGACCCGCAGCAGCGGCTCCTCCTCGAAACCTCCTGGGAG 3624

Qy        301 ACCATCGAACACGCCGGCATCAACCCCCACACCCTCCACGGCACCCCCACCGGAGTCTTC 360
              || ||||  |||||||  |||  ||||   | ||||  |  ||     |  |||  ||||||
Db       3625 GCCGTCGAGGACGCCGGGATCGACCCGACCTCCCTGCGGGGACGGCAGGTCGGCGTCTTC 3684

Qy        361 ACCGGCACCAACGGACAGGACTACGCACTTCGCGTGCACAACGCGGGCCAGTCAACCGAT 420
              || ||  | |     || || ||||  |   || |||   |||  ||  |     |||
Db       3685 ACTGGGGCGATGACCCACGAGTACGGGCCGAGCCTGCGGGACGGCGGGGAAGGCCTCGAC 3744

Qy        421 GGTTTCGCACTGACCGGAACCGCCGGCAGCGTCATCTCCGGTCGTATCTCGTACACGTTT 480
              || | |  |||||||| | | | ||||||  || ||  ||  ||  |||||||||  |
Db       3745 GGCTACCTGCTGACCGGCAACACGGCCAGCGTGATGTCGGGCCGCGTCTCGTACACACTC 3804

Qy        481 GGTTTTGAGGGTCCTGCGGTGTCGGTGGACACGGCTTGTTCCTCGTCGTTGGTGGCTTTG 540
              ||  |||||||  ||  ||  || ||||||||||||||  ||  || ||||||  |||| ||  ||
Db       3805 GGCCTTGAGGGCCCCGCCCTGACGGTGGACACGGCCTGCTCGTCGTCGCTGGTCGCCCTG 3864

Qy        541 CATCTGGCCTGTCAGGCGTTGCGTGCGGGTGAGTGCTCGATGGCGCTTGCCGGGGGTGTG 600
              || || |||   |||||  ||||   ||| |||   |  |||||||| |||||  || |||
Db       3865 CACCTCGCCGTGCAGGCCCTGCGCAAGGGCGAGGTCGACATGGCGCTCGCCGGCGGCGTG 3924

Qy        601 ACGGTGATGTCGTCTCCGGGTGCCTTCGTGGAGTTTTCGCGGCAGCGGGGTCTGGCCGCG 660
              | ||||||  |   | || ||     ||||| |||||     |||||||| || ||||||| |
Db       3925 GCCGTGATGCCCACGCCCGGGATGTTCGTCGAGTTCAGCCGGCAGCGCGGGCTGGCCGGG 3984

Qy        661 GACGGGCATTGCAAGGCGTTCTCGGCGGCGGCGGACGGGACCGGCTGGGGTGAGGGTGTG 720
              |||||  |   | |||||||| | ||| |||||||||| ||| |||||    ||||| ||
Db       3985 GACGGCCGGTCGAAGGCGTTCGCCGCGTCGGCGGACGGCACCAGCTGGTCCGAGGGCGTC 4044

Qy        721 GGGATGCTGCTGGTGGAGCGGCTCTCCGACGCCCATCGCAACGGTCACCGTGTCCTGGCC 780
              ||  | || || || ||||| || || |||||||  |||||||| ||||  |||||  ||
Db       4045 GGCGTCCTCCTCGTCGAGCGCCTGTCGGACGCCCGCCGCAACGGACACCAGGTCCTCGCG 4104

Qy        781 GTGGTGCGTGGCAGTGCGGTCAACCAGGACGGTGCGAGCAACGGTCTGACCGCGCCCAAC 840
              || || || ||||| ||  | |||||||||| |||||||||| || || || || |||
Db       4105 GTCGTCCGCGGCAGCGCCTTGAACCAGGACGGCGCGAGCAACGGCCTCACGGCTCCGAAC 4164

Qy        841 GGGCCGTCCCAGCAGCGTGTCATCCGCCAGGCCCTCGCCAACGCCGGCTTGTCGGCCGGT 900
              ||||| || |||||||| |||||||| |  || || ||  |||||  |  || || ||
Db       4165 GGGCCCTCGCAGCAGCGCGTCATCCGGCGCGCGCTGGCGGACGCCCGGCTGACGACCTCC 4224

Qy        901 GATGTCGACGCGGTGGAGGCCCACGGCACCGGCACCACTTTGGGCGACCCGATCGAGGCC 960
              || || ||||  || ||||| |||||||| ||||      | ||||||||||||||||||
Db       4225 GACGTGGACGTCGTCGAGGCACACGGCACGGGCACGCGACTCGGCGACCCGATCGAGGCG 4284

Qy        961 CAGGCCCTCCTCGCGACCTACGGACAGGACCGTGCCGGCGAGGGGCCGCTGTGGCTGGGC 1020
              ||||||||  |||| |||||||| |||| ||||| || |||   |||||||  | || ||
Db       4285 CAGGCCCTGATCGCCACCTACGGCCAGGGCCGTGACGACGAACAGCCGCTGCGCCTCGGG 4344

Qy       1021 TCGGTCAAGTCCAATGTCGGTCACACACAGGCTGCCGCGGGCGTCGCCGGGGTGATCAAG 1080
```

```
Db       4345 TCGTTGAAGTCCAACATCGGGCACACCCAGGCCGCGGCCGGCGTCTCCGGTGTCATCAAG 4404

Qy       1081 ATGGTGATGGCGCTGCGGCATGGTCTGCTGCCGCGGACGTTGCATGTGGATGAGCCGTCG 1140

Db       4405 ATGGTCCAGGCGATGCGCCACGGACTGCTGCCGAAGACGCTGCACGTCGACGAGCCCTCG 4464

Qy       1141 CCGCATGTGGACTGGTCCGCGGGTGCGGTGCAGCTGCTGACGGAGACGGTGCCCTGGCCC 1200

Db       4465 GACCAGATCGACTGGTCGGCTGGCGCCGTGGAACTCCTCACCGAGGCCGTCGACTGGCCG 4524

Qy       1201 G---GCGGGGAGGGGCGGCTACGGCGGGCAGGAGTGTCATCATTCGGCGTCAGCGGCACC 1257

Db       4525 GAGAAGCAGGACGGCGGGCTGCGCCGGGCCGCCGTCTCCTCCTTCGGGATCAGCGGCACC 4584

Qy       1258 AACGCCCACGTCATCCTCGAAGAAGCACCCG 1288

Db       4585 AATGCGCATGTGGTGCTCGAAGAGGCCCCGG 4615


RESULT 14
US-09-860-846-1
; Sequence 1, Application US/09860846
; Patent No. US20020164742A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/860,846
;  CURRENT FILING DATE:  2001-05-18
;  PRIOR APPLICATION NUMBER: 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 1
;   LENGTH: 15872
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-860-846-1

  Query Match             52.7%;  Score 680.4;  DB 9;  Length 15872;
  Best Local Similarity   71.0%;  Pred. No. 2.6e-156;
  Matches  916;  Conservative    0;  Mismatches  371;  Indels    3;  Gaps    1;

Qy          1 GAGCCGATCGCCATCGTTGGCATGGCCTGTCGTTTCCCCGGCGGAGTGACCTCGGCGGAC 60

Db       8498 GATCCGATCGCGATCGTGGGCATGGCATGCCGCTATCCGGGTGGTGTGTCGTCGCCGGAG 8557

Qy         61 GACTTCTGGGATCTGATCTCCTCCGAGCAGGACGCGATCGGCGGATTCCCCACCGACCGC 120

Db       8558 GACCTGTGGCGGCTGGTGGCCGAGGGGACGGACGCGATCAGCGAGTTCCCCGTCAACCGC 8617

Qy        121 GGCTGGGACCTGGACACGCTCTACGACCCCGACCCCGACCACCCCGGCACCTGCTACACC 180

Db       8618 GGCTGGGACCTGGAGAGCCTCTACGACCCGGATCCCGAGTCGAAGGGCACCACGTACTGC 8677

Qy        181 CGAAACGGCGGATTCCTCTACGACGCAGGCCACTTCGACGCCGAATTCTTCGGCATCAGC 240

Db       8678 CGGGAGGGCGGGTTCCTGGAAGGCGCCGGTGACTTCGACGCCGCCTTCTTCGGCATCTCG 8737

Qy        241 CCCCGCGAAGCCCTCGCCATGGACCCCCAGCAACGACTCCTCCTCGAAACCGCCTGGGAA 300

Db       8738 CCGCGCGAGGCCCTGGTGATGGACCCGCAGCAGCGGCTGCTGCTGGAGGTGTCCTGGGAG 8797

Qy        301 ACCATCGAACACGCCGGCATCAACCCCCACACCCTCCACGGCACCCCCACCGGAGTCTTC 360

Db       8798 GCGCTGGAACGCGCGGGCATCGACCCGTCCTCGCTGCGCGGCAGCCGCGGTGGTGTCTAC 8857

Qy        361 ACCGGCACCAACGGACAGGACTACGCACTTCGCGTGCACAACGCGGGCCAGTCAACCGAT 420

Db       8858 GTGGGCGCCGCGCACGGCTCGTACGCCTCCGATCCCCGGCTGGTGCCCGAGGGCTCGGAG 8917

Qy        421 GGTTTCGCACTGACCGGAACCGCCGGCAGCGTCATCTCCGGTCGTATCTCGTACACGTTT 480

Db       8918 GGCTATCTGCTGACCGGCAGCGCCGACGCGGTGATGTCCGGCCGCATCTCCTACGCGCTC 8977

Qy        481 GGTTTTGAGGGTCCTGCGGTGTCGGTGGACACGGCTTGTTCCTCGTCGTTGGTGGCTTTG 540

Db       8978 GGTCTCGAAGGACCGTCCATGACGGTGGAGACGGCCTGCTCCTCCTCGCTGGTGGCGCTG 9037

Qy        541 CATCTGGCCTGTCAGGCGTTGCGTGCGGGTGAGTGCTCGATGGCGCTTGCCGGGGGTGTG 600

Db       9038 CATCTGGCGGTACGGGCGCTGCGGCACGGCGAGTGCGGGCTCGCGCTGGCGGGCGGGGTG 9097

Qy        601 ACGGTGATGTCGTCTCCGGGTGCCTTCGTGGAGTTTTCGCGGCAGCGGGGTCTGGCCGCG 660

Db       9098 GCGGTGATGGCCGATCCGGCGGCGTTCGTGGAGTTCTCCCGGCAGAAGGGGCTGGCCGCC 9157

Qy        661 GACGGGCATTGCAAGGCGTTCTCGGCGGCGGCGGACGGGACCGGCTGGGGTGAGGGTGTG 720

Db       9158 GACGGCCGCTGCAAGGCGTTCTCGGCCGCCGCCGACGGCACCGGCTGGGCCGAGGGCGTC 9217

Qy        721 GGGATGCTGCTGGTGGAGCGGCTCTCCGACGCCCATCGCAACGGTCACCGTGTCCTGGCC 780

Db       9218 GGCGTGCTCGTCCTGGAGCGGCTGTCGGACGCGCGCCGCGCGGGGCACACGGTCCTCGGC 9277

Qy        781 GTGGTGCGTGGCAGTGCGGTCAACCAGGACGGTGCGAGCAACGGTCTGACCGCGCCCAAC 840

Db       9278 CTGGTCACCGGCACCGCGGTCAACCAGGACGGTGCCTCCAACGGGCTGACCGCGCCCAAC 9337

Qy        841 GGGCCGTCCCAGCAGCGTGTCATCCGCCAGGCCCTCGCCAACGCCGGCTTGTCGGCCGGT 900

Db       9338 GGCCCAGCCCAGCAACGCGTCATCGCCGAGGCGCTCGCCGACGCCGGGCTGTCCCCGGAG 9397

Qy        901 GATGTCGACGCGGTGGAGGCCCACGGCACCGGCACCACTTTGGGCGACCCGATCGAGGCC 960

Db       9398 GACGTGGACGCGGTCGAGGCGCACGGCACCGGCACCCGGCTCGGCGACCCCATCGAGGCC 9457

Qy        961 CAGGCCCTCCTCGCGACCTACGGACAGGACCGTGCCGGCGAGGGGCCGCTGTGGCTGGGC 1020

Db       9458 GGGGCGCTGCTCGCCGCCTCCGGACGGAACCGTTCCGGCGACCACCCGCTGTGGCTCGGC 9517

Qy       1021 TCGGTCAAGTCCAATGTCGGTCACACACAGGCTGCCGCGGGCGTCGCCGGGGTGATCAAG 1080

Db       9518 TCGCTGAAGTCCAACATCGGGCATGCCCAGGCCGCCGCCGGTGTCGGCGGCGTCATCAAG 9577

Qy       1081 ATGGTGATGGCGCTGCGGCATGGTCTGCTGCCGCGGACGTTGCATGTGGATGAGCCGTCG 1140

Db       9578 ATGCTCCAGGCGCTGCGGCACGGCTTGCTGCCCCGCACCCTCCACGCCGACGAGCCGACC 9637

Qy       1141 CCGCATGTGGACTGGTCCGCGGGTGCGGTGCAGCTGCTGACGGAGACGGTGCCCTGGCCC 1200

Db       9638 CCGCATGCCGACTGGAGCTCCGGCCGGGTACGGCTGCTCACCTCCGAGGTGCCGTGGC-- 9695

Qy       1201 GGCGGGGAGGGGCGGCTACGGCGGGCAGGAGTGTCATCATTCGGCGTCAGCGGCACCAAC 1260

Db       9696 -AGCGGACCGGCCGGCCCCGGCGGACCGGGGTGTCCGCCTTCGGCGTCGGCGGCACCAAT 9754

Qy       1261 GCCCACGTCATCCTCGAAGAAGCACCCGCC 1290

Db       9755 GCCCATGTCGTCCTCGAAGAGGCACCCGCC 9784


RESULT 15
US-09-988-384B-1
; Sequence 1, Application US/09988384B
; Publication No. US20030073824A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
```

```
;  FILE REFERENCE: 600.536US1
;  CURRENT APPLICATION NUMBER: US/09/988,384B
;  CURRENT FILING DATE:  2001-11-19
;  PRIOR APPLICATION NUMBER: PCT/US99/14398
;  PRIOR FILING DATE: 1999-06-25
;  PRIOR APPLICATION NUMBER: US 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 53
; SEQ ID NO 1
;   LENGTH: 15872
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-988-384B-1

  Query Match           52.7%;  Score 680.4;  DB 9;  Length 15872;
  Best Local Similarity  71.0%;  Pred. No. 2.6e-156;
  Matches  916;  Conservative    0;  Mismatches  371;  Indels    3;  Gaps    1;

Qy          1 GAGCCGATCGCCATCGTTGGCATGGCCTGTCGTTTCCCCGGCGGAGTGACCTCGGCGGAC 60
              || |||||||  ||||| |||||||| || || |  || || || ||| | ||| ||||
Db       8498 GATCCGATCGCGATCGTGGGCATGGCATGCCGCTATCCGGGTGGTGTGTCGTCGCCGGAG 8557

Qy         61 GACTTCTGGGATCTGATCTCCTCCGAGCAGGACGCGATCGGCGGATTCCCCACCGACCGC 120
              ||| | |||   ||| |  ||   | |  |||||||||| |||  ||||||  | |||||
Db       8558 GACCTGTGGCGGCTGGTGGCCGAGGGGACGGACGCGATCAGCGAGTTCCCCGTCAACCGC 8617

Qy        121 GGCTGGGACCTGGACACGCTCTACGACCCCGACCCCGACCACCCCGGCACCTGCTACACC 180
              |||||||||||||| |  |||||||||| || |||||        ||||||   |||  |
Db       8618 GGCTGGGACCTGGAGAGCCTCTACGACCCGGATCCCGAGTCGAAGGGCACCACGTACTGC 8677

Qy        181 CGAAACGGCGGATTCCTCTACGACGCAGGCCACTTCGACGCCGAATTCTTCGGCATCAGC 240
              ||  | ||||| |||||  | | ||| ||  ||||||||||||  ||||||||||||
Db       8678 CGGGAGGGCGGGTTCCTGGAAGGCGCCGGTGACTTCGACGCCGCCTTCTTCGGCATCTCG 8737

Qy        241 CCCCGCGAAGCCCTCGCCATGGACCCCCAGCAACGACTCCTCCTCGAAACCGCCTGGGAA 300
              || ||||| ||||| |  |||||||| ||||| || || || || ||      |||||||
Db       8738 CCGCGCGAGGCCCTGGTGATGGACCCGCAGCAGCGGCTGCTGCTGGAGGTGTCCTGGGAG 8797

Qy        301 ACCATCGAACACGCCGGCATCAACCCCCACACCCTCCACGGCACCCCCACCGGAGTCTTC 360
              |  | |||| ||| |||||| ||||   | | || | ||||| || |   || |||| |
Db       8798 GCGCTGGAACGCGCGGGCATCGACCCGTCCTCGCTGCGCGGCAGCCGCGGTGGTGTCTAC 8857

Qy        361 ACCGGCACCAACGGACAGGACTACGCACTTCGCGTGCACAACGCGGGCCAGTCAACCGAT 420
              |||  ||              |||||              |  | |  | ||    | ||
Db       8858 GTGGGCGCCGCGCACGGCTCGTACGCCTCCGATCCCCGGCTGGTGCCCGAGGGCTCGGAG 8917

Qy        421 GGTTTCGCACTGACCGGAACCGCCGGCAGCGTCATCTCCGGTCGTATCTCGTACACGTTT 480
              || |      |||||||| | |||||  |   || || ||||| || ||||| ||| || |
Db       8918 GGCTATCTGCTGACCGGCAGCGCCGACGCGGTGATGTCCGGCCGCATCTCCTACGCGCTC 8977

Qy        481 GGTTTTGAGGGTCCTGCGGTGTCGGTGGACACGGCTTGTTCCTCGTCGTTGGTGGCTTTG 540
              ||| | || || ||  |  || |||||||  ||||| || ||||| ||| |||||||  ||
Db       8978 GGTCTCGAAGGACCGTCCATGACGGTGGAGACGGCCTGCTCCTCCTCGCTGGTGGCGCTG 9037

Qy        541 CATCTGGCCTGTCAGGCGTTGCGTGCGGGTGAGTGCTCGATGGCGCTTGCCGGGGGTGTG 600
              ||||||||    | |||| ||||    || ||||||  | | ||||| || || || |||
Db       9038 CATCTGGCGGTACGGGCGCTGCGGCACGGCGAGTGCGGGCTCGCGCTGGCGGGCGGGGTG 9097

Qy        601 ACGGTGATGTCGTCTCCGGGTGCCTTCGTGGAGTTTTCGCGGCAGCGGGGTCTGGCCGCG 660
              ||||||||  |   |||||  || |||||||||||  || ||||||  ||| ||||||||
Db       9098 GCGGTGATGGCCGATCCGGCGGCGTTCGTGGAGTTCTCCCGGCAGAAGGGGCTGGCCGCC 9157

Qy        661 GACGGGCATTGCAAGGCGTTCTCGGCGGCGGCGGACGGGACCGGCTGGGGTGAGGGTGTG 720
              |||||  |  |||||||||||||||| || || ||||| |||||||||  |||||  ||
Db       9158 GACGGCCGCTGCAAGGCGTTCTCGGCCGCCGCCGACGGCACCGGCTGGGCCGAGGGCGTC 9217

Qy        721 GGGATGCTGCTGGTGGAGCGGCTCTCCGACGCCCATCGCAACGGTCACCGTGTCCTGGCC 780
              ||  ||||  |  |||||||||| || ||||| |  |||   || |||   |||| | |
Db       9218 GGCGTGCTCGTCCTGGAGCGGCTGTCGGACGCGCGCCGCGCGGGGCACACGGTCCTCGGC 9277

Qy        781 GTGGTGCGTGGCAGTGCGGTCAACCAGGACGGTGCGAGCAACGGTCTGACCGCGCCCAAC 840
              ||||    ||||  |||||||||||||||||||||    |||||| ||||||||||||||
Db       9278 CTGGTCACCGGCACCGCGGTCAACCAGGACGGTGCCTCCAACGGGCTGACCGCGCCCAAC 9337

Qy        841 GGGCCGTCCCAGCAGCGTGTCATCCGCCAGGCCCTCGCCAACGCCGGCTTGTCGGCCGGT 900
              || ||  ||||||| || ||||||  | |||| |||||| |||||||  ||||  | |
Db       9338 GGCCCAGCCCAGCAACGCGTCATCGCCGAGGCGCTCGCCGACGCCGGGCTGTCCCCGGAG 9397

Qy        901 GATGTCGACGCGGTGGAGGCCCACGGCACCGGCACCACTTTGGGCGACCCGATCGAGGCC 960
              || || |||||||| ||||| |||||||||||||||     | ||||||| |||||||||
Db       9398 GACGTGGACGCGGTCGAGGCGCACGGCACCGGCACCCGGCTCGGCGACCCCATCGAGGCC 9457

Qy        961 CAGGCCCTCCTCGCGACCTACGGACAGGACCGTGCCGGCGAGGGGCCGCTGTGGCTGGGC 1020
                ||| || |||||  ||| ||||| | |||||  |||||||     |||||||||| |||
Db       9458 GGGGCGCTGCTCGCCGCCTCCGGACGGAACCGTTCCGGCGACCACCCGCTGTGGCTCGGC 9517

Qy       1021 TCGGTCAAGTCCAATGTCGGTCACACACAGGCTGCCGCGGGCGTCGCCGGGGTGATCAAG 1080
              ||| |  ||||||||  |||| ||  | ||||| |||||  || ||||  ||| || ||||||
Db       9518 TCGCTGAAGTCCAACATCGGGCATGCCCAGGCCGCCGCCGGTGTCGGCGGCGTCATCAAG 9577

Qy       1081 ATGGTGATGGCGCTGCGGCATGGTCTGCTGCCGCGGACGTTGCATGTGGATGAGCCGTCG 1140
              |||  |  ||||||||||||  ||  |||||||  ||  ||  |  ||  |  || ||||||  |
Db       9578 ATGCTCCAGGCGCTGCGGCACGGCTTGCTGCCCCGCACCCTCCACGCCGACGAGCCGACC 9637

Qy       1141 CCGCATGTGGACTGGTCCGCGGGTGCGGTGCAGCTGCTGACGGAGACGGTGCCCTGGCCC 1200
              |||||||  ||||||  | | ||   ||| | ||||||  ||      ||||||  ||||
Db       9638 CCGCATGCCGACTGGAGCTCCGGCCGGGTACGGCTGCTCACCTCCGAGGTGCCGTGGC-- 9695

Qy       1201 GGCGGGGAGGGGCGGCTACGGCGGGCAGGAGTGTCATCATTCGGCGTCAGCGGCACCAAC 1260
                 ||   || ||||  ||||||  | ||  |||||  | |||||||| ||||||||||
Db       9696 -AGCGGACCGGCCGGCCCCGGCGGACCGGGGTGTCCGCCTTCGGCGTCGGCGGCACCAAT 9754

Qy       1261 GCCCACGTCATCCTCGAAGAAGCACCCGCC 1290
              ||||| ||| |||||||||| |||||||||
Db       9755 GCCCATGTCGTCCTCGAAGAGGCACCCGCC 9784


Search completed: June 18, 2003, 05:08:34
Job time : 159.541 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 17, 2003, 11:26:45 ; Search time 1653.65 Seconds
(without alignments)
12633.968 Million cell updates/sec

Title: US-09-914-286-1_COPY_6256_7545
Perfect score: 1290
Sequence: 1 gagccgatcgccatcgttgg.....tcctcgaagaagcacccgcc 1290

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
c	1	135.6	10.5	625	13	BM091001	BM091001 ig21a10.y
c	2	130.8	10.1	766	10	BE299394	BE299394 601118872
	3	123.8	9.6	627	13	BM091310	BM091310 ig21a10.x
	4	109.8	8.5	662	9	AI514012	AI514012 GH27052.5
c	5	108.4	8.4	800	17	AQ287101	AQ287101 mgxb0017L
	6	103.8	8.0	690	10	BB865539	BB865539 BB865539
	7	103.4	8.0	579	9	AI390803	AI390803 mb23h09.y
	8	101	7.8	640	10	BB866094	BB866094 BB866094
	9	100	7.8	682	10	BB866070	BB866070 BB866070
	10	99.4	7.7	546	10	AW659924	AW659924 97589 MAR
	11	98.4	7.6	638	10	BB866199	BB866199 BB866199
	12	96	7.4	681	10	BB865388	BB865388 BB865388
	13	95.8	7.4	701	10	BB865339	BB865339 BB865339
	14	95.2	7.4	664	10	BB865774	BB865774 BB865774
	15	94.6	7.3	516	9	AI532314	AI532314 SD03777.5
	16	94.2	7.3	663	10	BB866472	BB866472 BB866472
	17	93.8	7.3	667	10	BB866334	BB866334 BB866334
	18	93.8	7.3	672	10	BB865852	BB865852 BB865852
	19	93.6	7.3	655	10	BB865451	BB865451 BB865451
	20	93.6	7.3	702	10	BB865372	BB865372 BB865372
	21	93	7.2	674	10	BB865570	BB865570 BB865570
	22	92.6	7.2	673	10	BB866364	BB866364 BB866364
	23	92.6	7.2	674	10	BB865617	BB865617 BB865617
	24	92.6	7.2	680	10	BB866082	BB866082 BB866082
	25	92.4	7.2	708	10	BB865572	BB865572 BB865572
	26	92.2	7.1	672	10	BB865553	BB865553 BB865553
	27	92	7.1	677	10	BB866270	BB866270 BB866270
	28	91.6	7.1	675	10	BB866460	BB866460 BB866460
	29	91	7.1	685	10	BB865479	BB865479 BB865479
	30	90.2	7.0	647	10	BB866252	BB866252 BB866252
	31	89.4	6.9	706	10	BB865821	BB865821 BB865821
	32	89.2	6.9	502	13	BI570498	BI570498 RH02843.5
	33	89.2	6.9	662	10	BB866227	BB866227 BB866227
	34	88.2	6.8	638	10	BB866450	BB866450 BB866450
	35	87.2	6.8	657	10	BB865775	BB865775 BB865775
	36	87	6.7	677	10	BB865435	BB865435 BB865435
	37	87	6.7	681	10	BB865871	BB865871 BB865871
	38	85.6	6.6	680	10	BB865987	BB865987 BB865987
	39	84.8	6.6	651	10	BB865750	BB865750 BB865750
	40	84.4	6.5	677	10	BB866003	BB866003 BB866003
	41	84	6.5	689	13	BM579969	BM579969 170006872
	42	82.8	6.4	679	10	BB865515	BB865515 BB865515
	43	82.8	6.4	987	14	BQ963372	BQ963372 AGENCOURT
	44	82.4	6.4	626	10	BB866347	BB866347 BB866347
	45	81.6	6.3	591	9	AI108098	AI108098 GH06839.5

ALIGNMENTS

RESULT 1
BM091001/c
LOCUS BM091001 625 bp mRNA linear EST 20-NOV-2001
DEFINITION ig21a10.y1 Human Fetal Pancreas 1B Homo sapiens cDNA 5' similar to TR:Q16702 Q16702 FATTY ACID SYNTHASE ;, mRNA sequence.
ACCESSION BM091001
VERSION BM091001.1 GI:17019967
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 625)
AUTHORS Melton,D., Brown,J., Kenty,G., Permutt,A., Lee,C., Kaestner,K., Lemishka,I., Scearce,M., Brestelli,J., Gradwohl,G., Clifton,S., Hillier,L., Marra,M., Pape,D., Wylie,T., Martin,J., Blistain,A., Schmitt,A., Theising,B., Ritter,E., Ronko,I., Bennett,J., Cardenas ,M., Gibbons,M., McCann,R., Cole,R., Tsagareishvili,R., Williams,T. , Jackson,Y. and Bowers,Y.
TITLE Endocrine Pancreas Consortium
JOURNAL Unpublished (2000)
COMMENT Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue
Endocrine Pancreas Consortium
Harvard University, Howard Hughes Medical Institute
Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge, MA 02138
Tel: 617-495-1812
Fax: 617-495-8557
Email: dmelton@biohp.harvard.edu

```
            Library was constructed by Dr. Hiroshi Inoue DNA sequencing by:
            Washington University Genome Sequencing Center For information on
            obtaining a clone please contact: Dr. Hiroshi Inoue
            (hinoue@imgate.wustl.edu)
            Possible reversed clone: similarity on wrong strand
            High quality sequence stop: 485.
FEATURES              Location/Qualifiers
     source           1. .625
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /clone_lib="Human Fetal Pancreas 1B"
                      /tissue_type="Fetal Pancreas (4 Pooled Donors, 18 - 20
                      weeks, Stratagene #738023)"
                      /dev_stage="Fetal Pancreas"
                      /note="Vector: pBluescript SK(-); Site_1: NotI; Site_2:
                      XhoI; cDNA made by oligo-dT priming. Size-selected on
                      agarose gel.  Average insert size ~1kb. 5' XhoI site was
                      destroyed after directional cloning. Amplified once.
                      Contact information: Hiroshi Inoue, MD, Metabolism Div.
                      (Alan Permutt Lab), Washington University School of
                      Medicine, Box 8127, 660 South Euclid Ave., St. Louis, MO
                      63110 E-mail: hinoue@imgate.wustl.edu Tel: 314-362-1916,
                      Fax:314-747-2692."
BASE COUNT       98 a    190 c    220 g    117 t
ORIGIN

  Query Match          10.5%;  Score 135.6;  DB 13;  Length 625;
  Best Local Similarity   53.5%;  Pred. No. 2.5e-16;
  Matches  338;  Conservative    0;  Mismatches  279;  Indels   15;  Gaps    2;

Qy        506 TGGACACGGCTTGTTCCTCGTCGTTGGTGGCTTTGCATCTGGCCTGTCAGGCGTTGCGTG 565
              ||||||| ||||| |||||     || ||||  ||||    ||||  |||||  | ||
Db        618 TGGACACAGCTTGCTCCTCCAGCCTGATGGCCCTGCAGAACGCCTACCAGGCCATCCACA 559

Qy        566 CGGGTGAGTGCTCGATGGCGCTTGCCGGGGGTGTGACGGTGATGTCGTCTCCGGGTGCCT 625
              ||  ||||| |    ||  | |  ||||  | |  ||  ||  |  ||     |||
Db        558 GCGGGCAGTGCCCTGCCGCCATCGTGGGGGGCATCAACGTCCTGCTGAAGCCCAACACCT 499

Qy        626 TCGTGGAGTTTTCGCGGCAGCGGGGTCTGGCCGCGGACGGGCATTGCAAGGCGTTCTCGG 685
              ||||  |||| | | ||| | ||   ||   | | || ||    |||||||| |||
Db        498 CCGTGCAGTTCTTGAGGCTGGGGATGCTCAGCCCCGAGGGCACCTGCAAGGCCTTCGACA 439

Qy        686 CGGCGGCGGACGGGACCGGCTGGGGTGAGGGTGTGGGGATGCTGCTGGTGGAGCGGCTCT 745
              | |||| | | |||  | || |    |||||||||| |    | ||| ||     |   |
Db        438 CAGCGGGGAATGGGTACTGCCGCTCGGAGGGTGTGGTGGCCGTCCTGCTGACCAAGAAGT 379

Qy        746 CCGACGCCCATCGCAACGGTCACCGTGTCCTGGCCGTGGTGCGTGGCAGTGCGGTCAACC 805
              ||   ||||  ||         |    |||||  ||  |        |  |  | |  |
Db        378 CCCTGGCCCGGCGGGTGTACGCCACCATCCTGAACGCCGGCACCAATACAGATGGCTTCA 319

Qy        806 AGGACGGTGCGAGCAACGGTCTGACCGCGCCCAACGGGCCGTCCCAGCAGCGTGTCATCC 865
              ||||          | ||  |||||   |||   |||     |||| |||   ||||||
Db        318 AGGA---------GCAAGGCGTGACCTTCCCCTCAGGGGATATCCAGGAGCAGCTCATCC 268

Qy        866 GCCAGGCCCTCGCCAACGCCGGCTTGTCGGCCGGTGATGTCGACGCGGTGGAGGCCCACG 925
              ||  |     |      |||||  || |  | |       | ||     | || |||||||
Db        267 GCTCGTTGTACCAGTCGGCCGGAGTGGCCCCTGAGTCATTTGAATACATCGAAGCCCACG 208

Qy        926 GCACCGGCACCACTTTGGGCGACCCGATCGAGGCCCAGGCCCTCCTCGCGACCTACGGAC 985
              |||| |||||||   |||||||||||     |||    | | | ||      |||   |  |
Db        207 GCACAGGCACCAAGGTGGGCGACCCCCAGGAGCTGAATGGCATC------ACCCGAGCCC 154

Qy        986 AGGACCGTGCCGGCGAGGGGCCGCTGTGGCTGGGCTCGGTCAAGTCCAATGTCGGTCACA 1045
              |  |    || || ||| |||||||    | |||||   |||||||||  | || |||
Db        153 TGTGCGCCACCCGCCAGGAGCCGCTGCTCATCGGCTCCACCAAGTCCAACATGGGGCACC 94

Qy       1046 CACAGGCTGCCGCGGGCGTCGCCGGGGTGATCAAGATGGTGATGGCGCTGCGGCATGGTC 1105
              |  || | ||| ||||   | || |    ||  |||| || || | | |||  ||| || |
Db         93 CGGAGCCAGCCTCGGGGCTGGCAGCCCTGGCCAAGGTGCTGCTGTCCCTGGAGCACGGGC 34

Qy       1106 TGCTGCCGCGGACGTTGCATGTGGATGAGCCG 1137
              |  | | |  |   |||| |  ||   |||
Db         33 TCTGGGCCCCCAACCTGCACTTCCATAGCCCG 2
```

```
RESULT 2
BE299394/c
LOCUS       BE299394                766 bp    mRNA    linear   EST 20-JUL-2000
DEFINITION  601118872F1 NIH_MGC_17 Homo sapiens cDNA clone IMAGE:3028841 5',
            mRNA sequence.
ACCESSION   BE299394
VERSION     BE299394.1  GI:9183142
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 766)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Ling Hong/Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
            Plate: LLCM87  row: o  column: 18
            High quality sequence start: 6
            High quality sequence stop: 674.
FEATURES              Location/Qualifiers
     source           1. .766
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /clone="IMAGE:3028841"
                      /clone_lib="NIH_MGC_17"
                      /tissue_type="rhabdomyosarcoma"
                      /lab_host="DH10B (phage-resistant)"
                      /note="Organ: muscle; Vector: pOTB7; Site_1: EcoRI;
                      Site_2: XhoI; cDNA made by oligo-dT priming.
                      Directionally cloned into EcoRI/XhoI sites using the
                      following 5' adaptor: GGCACGAG(G).  Size-selected >500bp
                      for average insert size 1.8kb.  Library constructed by
                      Ling Hong in the laboratory of Gerald M. Rubin (University
                      of California, Berkeley) using ZAP-cDNA synthesis kit
                      (Stratagene) and Superscript II RT (Life Technologies)."
BASE COUNT      135 a    231 c    248 g    152 t
ORIGIN

  Query Match          10.1%;  Score 130.8;  DB 10;  Length 766;
  Best Local Similarity   50.9%;  Pred. No. 2.2e-15;
  Matches  342;  Conservative    0;  Mismatches  322;  Indels    8;  Gaps    1;

Qy        241 CCCCGCGAAGCCCTCGCCATGGACCCCCAGCAACGACTCCTCCTCGAAACCGCCTGGGAA 300
              ||||  | |||      |  |||| ||||  || || || || |||  | |||  |||
Db        674 CCCCAAGCAGCAACACACGATGACCCTCAGCTGCGGCTGCTGCTGGAAGTCACCTATGAA 615

Qy        301 ACCATCGAACACGCCGGCATCAACCCCCACACCCTCCACGGCACCCCCACCGGAGTCTTC 360
              ||||||   ||| ||||||||||||  |  | ||||  || || | ||| || ||||
Db        614 GCCATCGTGGACGGAGGCATCAACCCAGATTCACTCCGAGGAACACACACTGGCGTCTGG 555

Qy        361 ACCGGCACCAACGGACAGGACTACGCACTTCGCGTGCACAACGCGGGCCAGTCAACCGAT 420
              |||   | |||     ||   | |       | ||  |    |     | || ||  ||
Db        554 GTGGGCGTGAGCGGCTCTGAGACCTCGGAGGCCCTGAGCCGAGACCCCGAGACACTCGTG 495

Qy        421 GGTTTCGCACTGACCGGAACCGCCGGCAGCGTCATCTCCGGTCGTATCTCGTACACGTTT 480
              || | |    ||  ||   |    |  | || ||   ||  |||| | |   ||
Db        494 GGCTACAGCATGGTGGGCTGCCAGCGAGCGATGATGGCCAACCGGCTCTCCTTCTTCTTC 435

Qy        481 GGTTTTGAGGGTCCTGCGGTGTCGGTGGACACGGCTTGTTCCTCGTCGTTGGTGGCTTTG 540
              |  ||    || ||     |  | ||||||| || || |||||     || ||||  ||
Db        434 GACTTCAGAGGGCCCAGCATCGCACTGGACACAGCCTGCTCCTCCAGCCTGATGGCCCTG 375

Qy        541 CATCTGGCCTGTCAGGCGTTGCGTGCGGGTGAGTGCTCGATGGCGCTTGCCGGGGGTGTG 600
```

```
Db        374 CAGAACGCCTACCAGGCCATCCACAGCGGGCAGTGCCCTGCCGCCATCGTGGGGGGCATC 315

Qy        601 ACGGTGATGTCGTCTCCGGGTGCCTTCGTGGAGTTTTCGCGGCAGCGGGGTCTGGCCGCG 660

Db        314 AACGTCCTGCTGAAGCCCAACACCTCCGTGCAGTTCTTGAGGCTGGGGATGCTCAGCCCC 255

Qy        661 GACGGGCATTGCAAGGCGTTCTCGGCGGCGGCGGACGGGACCGGCTGGGGTGAGGGTGTG 720

Db        254 GAGGGCACCTGCAAGGCCTTCGACACAGCGGGGAATGGGTACTGCCGCTCGGAGGGTGTG 195

Qy        721 GGGATGCTGCTGGTGGAGCGGCTCTCCGACGCCCATCGCAACGGTCACCGTGTCCTGGCC 780

Db        194 GTGGCCGTCCTGCTGACCAAGAAGTCCCTGGCCCGGCGGGTGTACGCCACCATCCTGAAC 135

Qy        781 GTGGTGCGTGGCAGTGCGGTCAACCAGGACGGTGCGAGCAACGGTCTGACCGCGCCCAAC 840

Db        134 GCCGGCACCAATACAGATGGCTTCAAGGAAGC--------AAGGCGTGACCTTCCCCTCA 83

Qy        841 GGGCCGTCCCAGCAGCGTGTCATCCGCCAGGCCCTCGCCAACGCCGGCTTGTCGGCCGGT 900

Db         82 GGGGATATCCAGGAGCAGCTCATCCGCTCGTTGTACCAGTCGGCCGGAGTGGCCCTGAGT 23

Qy        901 GATGTCGACGCG 912

Db         22 CATTTCCTCGTG 11
```

RESULT 3
BM091310
LOCUS BM091310 627 bp mRNA linear EST 20-NOV-2001
DEFINITION ig21a10.x1 Human Fetal Pancreas 1B Homo sapiens cDNA 3' similar to TR:Q16702 Q16702 FATTY ACID SYNTHASE ;, mRNA sequence.
ACCESSION BM091310
VERSION BM091310.1 GI:17020276
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 627)
AUTHORS Melton,D., Brown,J., Kenty,G., Permutt,A., Lee,C., Kaestner,K., Lemishka,I., Scearce,M., Brestelli,J., Gradwohl,G., Clifton,S., Hillier,L., Marra,M., Pape,D., Wylie,T., Martin,J., Blistain,A., Schmitt,A., Theising,B., Ritter,E., Ronko,I., Bennett,J., Cardenas,M., Gibbons,M., McCann,R., Cole,R., Tsagareishvili,R., Williams,T., Jackson,Y. and Bowers,Y.
TITLE Endocrine Pancreas Consortium
JOURNAL Unpublished (2000)
COMMENT Other_ESTs: ig21a10.y1
Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue
Endocrine Pancreas Consortium
Harvard University, Howard Hughes Medical Institute
Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge, MA 02138
Tel: 617-495-1812
Fax: 617-495-8557
Email: dmelton@biohp.harvard.edu
Library was constructed by Dr. Hiroshi Inoue DNA sequencing by: Washington University Genome Sequencing Center For information on obtaining a clone please contact: Dr. Hiroshi Inoue (hinoue@imgate.wustl.edu)
Possible reversed clone: similarity on wrong strand
High quality sequence stop: 480.
FEATURES Location/Qualifiers
source 1. .627
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="Human Fetal Pancreas 1B"
/tissue_type="Fetal Pancreas (4 Pooled Donors, 18 - 20 weeks, Stratagene #738023)"
/dev_stage="Fetal Pancreas"
/note="Vector: pBluescript SK(-); Site_1: NotI; Site_2: XhoI; cDNA made by oligo-dT priming. Size-selected on agarose gel. Average insert size ~1kb. 5' XhoI site was destroyed after directional cloning. Amplified once. Contact information: Hiroshi Inoue, MD, Metabolism Div. (Alan Permutt Lab), Washington University School of Medicine, Box 8127, 660 South Euclid Ave., St. Louis, MO 63110 E-mail: hinoue@imgate.wustl.edu Tel: 314-362-1916, Fax:314-747-2692."
BASE COUNT 119 a 206 c 196 g 106 t
ORIGIN

Query Match 9.6%; Score 123.8; DB 13; Length 627;
Best Local Similarity 53.1%; Pred. No. 4.7e-14;
Matches 263; Conservative 0; Mismatches 232; Indels 0; Gaps 0;

```
Qy        264 CCCCCAGCAACGACTCCTCCTCGAAACCGCCTGGGAAACCATCGAACACGCCGGCATCAA 323

Db         39 CCCTCAGCTGCGGCTGCTGCTGGAAGTCACCTATGAAGCCATCGTGGACGGAGGCATCAA 98

Qy        324 CCCCCACACCCTCCACGGCACCCCCACCGGAGTCTTCACCGGCACCAACGGACAGGACTA 383

Db         99 CCCAGATTCACTCCGAGGAACACACACTGGCGTCTGGGTGGGCGTGAGCGGCTCTGAGAC 158

Qy        384 CGCACTTCGCGTGCACAACGCGGGCCAGTCAACCGATGGTTTCGCACTGACCGGAACCGC 443

Db        159 CTCGGAGGCCCTGAGCCGAGACCCCGAGACACTCGTGGGCTACAGCATGGTGGGCTGCCA 218

Qy        444 CGGCAGCGTCATCTCCGGTCGTATCTCGTACACGTTTGGTTTTGAGGGTCCTGCGGTGTC 503

Db        219 GCGAGCGATGATGGCCAACCGGCTCTCCTTCTTCTTCGACTTCAGAGGGCCCAGCATCGC 278

Qy        504 GGTGGACACGGCTTGTTCCTCGTCGTTGGTGGCTTTGCATCTGGCCTGTCAGGCGTTGCG 563

Db        279 ACTGGACACAGCCTGCTCCTCCAGCCTGATGGCCCTGCAGAACGCCTACCAGGCCATCCA 338

Qy        564 TGCGGGTGAGTGCTCGATGGCGCTTGCCGGGGGGTGTGACGGTGATGTCGTCTCCGGGTGC 623

Db        339 CAGCGGGCAGTGCCCTGCCGCCATCGTGGGGGGCATCAACGTCCTGCTGAAGCCCAACAC 398

Qy        624 CTTCGTGGAGTTTTCGCGGCAGCGGGGTCTGGCCGCGGACGGGCATTGCAAGGCGTTCTC 683

Db        399 CTCCGTGCAGTTCTTGAGGCTGGGGATGCTCAGCCCCGAGGGCACCTGCAAGGCCTTCGA 458

Qy        684 GGCGGCGGCGGACGGGACCGGCTGGGGTGAGGGTGTGGGGATGCTGCTGGTGGAGCGGCT 743

Db        459 CACAGCGGGGAATGGGTACTGCCGCTCGGAGGGTGTGGTGGCCGTCCTGCTGACCAAGAA 518

Qy        744 CTCCGACGCCCATCG 758

Db        519 GTCCCTGGGCCGGCG 533
```

RESULT 4
AI514012
LOCUS AI514012 662 bp mRNA linear EST 23-APR-2001
DEFINITION GH27052.5prime GH Drosophila melanogaster head pOT2 Drosophila melanogaster cDNA clone GH27052 5 similar to BcDNA:GH07626: FBan0003523 located on: 2L 23D1-23D1;: 04/10/2001, mRNA sequence.
ACCESSION AI514012
VERSION AI514012.1 GI:4418074
KEYWORDS EST.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 662)
AUTHORS Harvey,D., Brokstein,P., Hong,L., Evans-Holm,M., Su,C., Tsang,G., Lewis,S. and Rubin,G.M.
TITLE BDGP/HHMI Drosophila EST Project
JOURNAL Unpublished (2001)
COMMENT Other_ESTs: GH27052.3prime
Contact: Stapleton, M.

```
            BDGP
            Lawrence Berkeley National Lab
            One Cyclotron Rd, Berkeley, CA 94720, USA
            Fax: 510 486 6798
            Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
            hit genomic AE003581: arm:2L [2747932,3054367]
            estimated-cyto:23B3-23D3: 04/10/2001
            Plate: GH.270  row: E  column: 4
            High quality sequence stop: 542
            POLYA=No.
FEATURES              Location/Qualifiers
     source           1. .662
                      /organism="Drosophila melanogaster"
                      /db_xref="taxon:7227"
                      /clone="GH27052"
                      /clone_lib="GH Drosophila melanogaster head pOT2"
                      /sex="male and female"
                      /dev_stage="adult"
                      /lab_host="DH5 - alpha"
                      /note="Organ: head; Vector: pOT2; Site_1: EcoRI; Site_2:
                      XhoI; Sized fractionated cDNAs were directly ligated into
                      pOT2.  Plasmid cDNA library."
BASE COUNT      143 a    185 c    185 g    149 t
ORIGIN

  Query Match             8.5%;  Score 109.8;  DB 9;  Length 662;
  Best Local Similarity   50.5%;  Pred. No. 2.3e-11;
  Matches  329;  Conservative    0;  Mismatches  307;  Indels   15;  Gaps    2;

Qy        515 CTTGTTCCTCGTCGTTGGTGGCTTTGCATCTGGCCTGTCAGGCGTTGCGTGCGGGTGAGT 574
              |||| |||   ||  ||   || ||| | | ||| |     |   |||| |  ||  ||
Db          1 CTTGCTCCAGTTCTCTGTACGCCTTGGAACAGGCTTTCTCCGATATGCGCGAAGGAAAGG 60

Qy        575 GCTCGATGGCGCTTGCCGGGGGTGTGACGGTGATGTCGTCTCCGGGTGCCTTCGTGGAGT 634
              |   | || || | ||  || |      ||||         ||        |  || |||
Db         61 TCGACAACGCCCTGGTCGCTGGAGCTGGACTGATCCTCAAGCCCACCATGTCGCTGCAGT 120

Qy        635 TTTCGCGGCAGCGGGGTCTGGCCGCGGACGGGCATTGCAAGGCGTTCTCGGCGGCGGCGG 694
              |   ||| | |        ||  | |||||||    |||||||| |||   | | | |
Db        121 TCAAGCGACTGAATATGTTGAGCCCGGACGGCAGCTGCAAGGCCTTCGATGAGTCTGGCA 180

Qy        695 ACGGGACCGGCTGGGGTGAGGGTGTGGGGATGCTGCTGGTGGAGCGGCTCTCCGACGCCC 754
              | ||   || | |   || ||    | | |||||||  || ||||   ||| |  |||
Db        181 ATGGATACGTCCGTTCCGATGGATGTGTGGTGCTGCTCCTGCAGCGCACCTCTGCAGCCA 240

Qy        755 ATCGCAACGGTCACCGTGTCCTGGCCGTGGTGCGTGGCAGTGCGGTCAACCAGGACGGTG 814
              ||        | |    ||||    |||          | |  |    | ||||
Db        241 GGCGTGTGTATGCCTCCATCCTCAATGTGCGCACCAATACGGATGGTTTCAAGGAGC--- 297

Qy        815 CGAGCAACGGTCTGACCGCGCCCAACGGGCCGTCCCAGCAGCGTGTCATCCGCCAGGCCC 874
                | ||  | ||    || |  ||    |   ||  | ||  | |||||| || ||
Db        298 ---,--AGGGCATCACATACCCTATTGGCAAGATGCAAAATCGCCTGATCCGCGAGACCT 351

Qy        875 TCGCCAACGCCGGCTTGTCGGCCGGTGATGTCGACGCGGTGGAGGCCCACGGCACCGGCA 934
              ||   |   ||  |    |||  ||||| |    |||||||| ||||| ||||| |
Db        352 ACGAGGAGATTGGTCTTAACCCCGCCGATGTGGTTTACGTGGAGGCACACGGTACCGGAA 411

Qy        935 CCACTTTGGGCGACCCGATCGAGGCCCAGGCCCTCCTCGCGACCTACGGACAGGACCGTG 994
              |||   |||||| ||    ||||   |  |  ||    |    || | |  |||||||
Db        412 CCAAGGTGGGCGATCCCCAGGAGGTGAACTCTATCACTGACTTCTTCTGCAAGGACCGTA 471

Qy        995 CCGGCGAGGGGCCGCTGTGGCTGGGCTCGGTCAAGTCCAATGTCGGTCACACACAGGCTG 1054
              |   |      || |||| | | || |||||||||||||||  | ||||||  |  | | |
Db        472 CGACC------CCTCTGTTGATCGGATCGGTCAAGTCCAACATGGGTCACTCGGAACCCG 525

Qy       1055 CCGCGGGCGTCGCCGGGGTGATCAAGATGGTGATGGCGCTGCGGCATGGTCTGCTGCCGC 1114
              |  | || |||  |   |||  ||||||  |||| ||  ||  | | ||  |  | ||
Db        526 CTTCCGGTGTCTGCTCTGTGGCCAAGATTCTGATCGCCATGGAGGAGGGCGTCATTCCCG 585

Qy       1115 GGACGTTGCATGTGGATGAGCCGTCGCCGCATGTGGACTGGTCCGCGGGTG 1165
              | |  |||||      |  |||||   ||  || | |  |   ||  | ||
Db        586 GTAACTTGCACTACAACAAGCCGAACCCAGATCTTTATGGACTCGTTGATG 636
```

```
RESULT 5
AQ287101/c
LOCUS       AQ287101                800 bp    DNA     linear   GSS 03-DEC-1998
DEFINITION  mgxb0017L22r CUGI Rice Blast BAC Library Magnaporthe grisea genomic
            clone mgxb0017L22r, DNA sequence.
ACCESSION   AQ287101
VERSION     AQ287101.1  GI:3948187
KEYWORDS    GSS.
SOURCE      Magnaporthe grisea.
  ORGANISM  Magnaporthe grisea
            Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
            Sordariomycetes incertae sedis; Magnaporthaceae; Magnaporthe.
REFERENCE   1  (bases 1 to 800)
  AUTHORS   Yu,Y., Zhu,H., Boyd,C.A., Gaudette,B., Gayle,A., Kingsbury,R.,
            Phillips,K., Sasinowski,M, Wing,R.A. and Dean,R.A.
  TITLE     A BAC End Sequencing Framework to Sequence the Magnaporthe grisea
            Genome
  JOURNAL   Unpublished (1998)
COMMENT     Contact: Dean RA
            Clemson University Genomics Institute
            Clemson University
            100 Jordan Hall, Clemson Universiy, Clemson, SC 29634
            Tel: 864 656 5737
            Fax: 864 656 4293
            Email: rdean@clemson.edu
            Seq primer: GGAAACAGCTATGACCATG
            Class: BAC ends
            High quality sequence start: 40
            High quality sequence stop: 512.
FEATURES              Location/Qualifiers
     source           1. .800
                      /organism="Magnaporthe grisea"
                      /strain="70-15"
                      /db_xref="taxon:148305"
                      /clone="mgxb0017L22r"
                      /clone_lib="CUGI Rice Blast BAC Library"
                      /tissue_type="Protoplasts"
                      /lab_host="E. coli DH10B"
                      /note="Vector: pBACWICH; Site_1: HindIII; Site_2: HindIII;
                      Rice blast is one of the most devestating fungal diseases
                      of rice  world wide.  It is a filamentous ascomycete with
                      a haploid genome (n=7)  of approximately 40 Mbp.  Rice
                      blast is an important model fungal pathogen  for studying
                      numerous aspects of the fungal-host interaction.   In
                      order to facilitate genome wide analysis, a BAC library
                      containing  9216 clones with an average insert size of 130
                      kbp was constructed.   This library represents greater
                      than 25X genome coverage.  High density  colony filters
                      are available upon request."
BASE COUNT      181 a    202 c    226 g    187 t      4 others
ORIGIN

  Query Match             8.4%;  Score 108.4;  DB 17;  Length 800;
  Best Local Similarity   51.2%;  Pred. No. 4.4e-11;
  Matches  303;  Conservative    0;  Mismatches  283;  Indels    6;  Gaps    2;

Qy          1 GAGCCGATCGCCATCGTTGGCATGGCCTGTCGTTTCCCCGGCGGAGTGACCTCGGCGGAC 60
              ||||| ||||||||| ||||||  | ||| || ||| ||||  |       |    |
Db        780 GAGCCTATCGCCATCATTGGCACCGGCTGCCGCTTCNCCGGAAGTTGCGACAGCCCCTCG 721

Qy         61 GACTTCTGGGATCTGATCTCCTCCGAGCAGGACGCGATCGGCGGATTCCCCACCGACCGC 120
              |  |||||||  || |       |  |||  | |     | |||| |  || ||
Db        720 AAGCTCTGGGAGTTGCTGGAAAAGCCACGCGACCTGTTGAAGAAAATCCCAAATGATCGA 661

Qy        121 GGCTGGGACCTGGACACGCTCTACGACCCCGACCCCGACCACCCCGGCACCTGCTACACC 180
                | |     ||   | | ||| |||| |   |   ||||  |||||  || ||
Db        660 TTTAGTG---CTGATGGGTTTTACCACCCAAAGAACTCTCACCATGGCACTAGCAACGTG 604

Qy        181 CGAAACGGCGGATTCCTCTACGACGCAGGCCA---CTTCGACGCCGAATTCTTCGGCATC 237
              ||| ||       | ||| |||| |      |   || ||| ||| | ||||| ||  ||
Db        603 CGACACTCTTACCTGCTCGACGAGGATNTGCGTACCTACGATGCCCAGTTCTTTGGGGTC 544
```

```
Qy        238 AGCCCCCGCGAAGCCCTCGCCATGGACCCCCAGCAACGACTCCTCCTCGAAACCGCCTGG 297
              |    |   ||| |||  | ||||||| || ||||| || || ||| | |||||||  |
Db        543 AAGGCGATCGAGGCCAGCTCCATGGATCCTCAGCAGCGTCTGCTCATGGAAACCGTGTAC 484

Qy        298 GAAACCATCGAACACGCCGGCATCAACCCCCACACCCTCCACGGCACCCCCACCGGAGTC 357
              ||  |  | |||   |||||  |   |  | |     || || ||| ||  ||| |  ||
Db        483 GAGGCTCTGGAATCGGCCGGGCTTTCCATCAAGCAACTTCAGGGCTCCGACACGGCCGTT 424

Qy        358 TTCACCGGCACCAACGGACAGGACTACGCACTTCGCGTGCACAACGCGGGCCAGTCAACC 417
              |      ||    |    ||  |       |       |       ||  | | |      |
Db        423 TATGTTGGTGTTATGAGTGCCGATTTTATGGATATGCTCGCGCGCGACGTCGAAAAGTTC 364

Qy        418 GATGGTTTCGCACTGACCGGAACCGCCGGCAGCGTCATCTCCGGTCGTATCTCGTACACG 477
                || |      || ||||| ||  ||||| |  |    |   ||  |||| |||
Db        363 CCAACTTACTTCGCCACTGGAACTGCACGCAGCATTTTGAGCAACCGCCTCTCCTACTTT 304

Qy        478 TTTGGTTTTGAGGGTCCTGCGGTGTCGGTGGACACGGCTTGTTCCTCGTCGTTGGTGGCT 537
              |  | ||   | |||||  |  || |  | ||||| || || || |||   ||| | ||
Db        303 TACGATTGGCACGGTCCATCCATGACAATCGACACAGCCTGCTCGTCGAGTTTGATTGCC 244

Qy        538 TTGCATCTGGCCTGTCAGGCGTTGCGTGCGGGTGAGTGCTCGATGGCGCTTG 589
              |||||||  ||    ||||  | ||| |     ||| |  | ||||  | |
Db        243 TTGCATCAAGCTGTACAGGTCCTTCGTTCCAAGCAGTCCAAGGTGGCAGTGG 192
```

RESULT 6
BB865539
LOCUS BB865539 690 bp mRNA linear EST 27-NOV-2001
DEFINITION BB865539 RIKEN full-length enriched, RCB-0545 OHTA cDNA Mus musculus cDNA clone G431001N10 5', mRNA sequence.
ACCESSION BB865539
VERSION BB865539.1 GI:17111749
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 690)
AUTHORS Akimura,T., Arakawa,T., Carninci,P., Furuno,M., Hanagaki,T., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T., Imotani,K., Ishii,Y., Ito,M., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Matsuyama,T., Nakamura,M., Nishi,K., Nomura,K., Numasaki,R., Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sakazume,N., Sasaki,D., Sato,K., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagawa,A., Takahashi,F., Takaku-Akahira,S., Tanaka,T., Tomaru,A., Toya,T., Watahiki,A., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y.
TITLE RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura,T., et al. 2001)
JOURNAL Unpublished (2001)
COMMENT Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.go.jp,
URL:http://genome.gsc.riken.go.jp/
Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. . 10 (10), 1617-1630 (2000)
wagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. . 10 (11), 1757-1771 (2000)
Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara,Y. and Hayashizaki,Y.
Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. . 11 (2), 281-289 (2001)
Please visit our web site (http://genome.gsc.riken.go.jp) for further details.
e mouse tissues.
FEATURES Location/Qualifiers
source 1. .690
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="G431001N10"
/clone_lib="RIKEN full-length enriched, RCB-0545 OHTA cDNA"
/cell_line="RCB-0545 OHTA"
/note="pooled cell lines ; (cell_line=CRL-1751 WEHI 164), (cell_line=CRL-2116 JC), (cell_line=RCB-0035 WEHI-3), (cell_line=RCB-0464 Meth-A), (cell_line=RCB-0545 OHTA), (cell_line=RCB-0559 K-1 . F1), (cell_line=RCB-1283 B16 melanoma), (cell_type=B cells, cell_line=CRL-1702 WEHI 231), (cell_type=Leydig cells, cell_line=CRL-2065 MLTC-1), (cell_type=Nullipotent stem cell, cell_line=CRL-2070 NE), (tissue_type=bladder, cell_line=RCB-0544 MBT-2), (tissue_type=bone marrow, cell_type=stroma cell, cell_line= CRL-2028 SR-4987), (tissue_type=colon, cell_line=RCB-0549 Cle-H3), (tissue_type=kidney, cell_line=CCL-142 RAG), (tissue_type=submandibular gland, cell_line=CRL-1734 SCA-9 clone 15), (strain=BALB/C, cell_type=B cells, cell_line=CRL-1669 BCL1 Clone 13.20-3B3), (strain=C3H, tissue_type=brain, cell_line=CRL-1443 BC3H1)"
BASE COUNT 146 a 190 c 209 g 145 t
ORIGIN

Query Match 8.0%; Score 103.8; DB 10; Length 690;
Best Local Similarity 51.2%; Pred. No. 3.3e-10;
Matches 243; Conservative 0; Mismatches 232; Indels 0; Gaps 0;

```
Qy        190 GGATTCCTCTACGACGCAGGCCACTTCGACGCCGAATTCTTCGGCATCAGCCCCCGCGAA 249
              |||   ||  | ||     | | |||||||||   || || ||  ||  ||||    |
Db        216 GGAAAGCTGAAGGATCTCTCCAAGTTCGACGCCTCCTTTTTTGGGGTCCACCCCAAGCAG 275

Qy        250 GCCCTCGCCATGGACCCCCAGCAACGACTCCTCCTCGAAACCGCCTGGGAAACCATCGAA 309
              || | | | ||||||||||||||  || || ||  | |||  |  ||  ||| | || |
Db        276 GCACACACAATGGACCCCCAGCTTCGGCTGCTGTTGGAAGTCAGCTATGAAGCAATTGTG 335

Qy        310 CACGCCGGCATCAACCCCCACACCCTCCACGGCACCCCCACCGGAGTCTTCACCGGCACC 369
              | |  || |||||||    | | ||||  || ||   ||| || ||||      ||
Db        336 GATGGAGGTATCAACCCAGCCTCACTCCGAGGAACGAACACTGGCGTCTGGGTGGGTGTG 395

Qy        370 AACGGACAGGACTACGCACTTCGCGTGCACAACGCGGGCCAGTCAACCGATGGTTTCGCA 429
              | ||   ||      |  | |   ||  |   | ||  |       || | |
Db        396 AGTGGTTCAGAGGCATCCGAGGCCCTTAGCAGAGATCCCGAGACGCTTCTGGGCTACAGC 455

Qy        430 CTGACCGGAACCGCCGGCAGCGTCATCTCCGGTCGTATCTCGTACACGTTTGGTTTTGAG 489
              ||   ||   |    |      | ||  ||   ||  |||| | |   || |  ||  |
Db        456 ATGGTGGGCTGCCAGCGTGCAATGATGGCCAACCGGCTCTCTTTCTTCTTCGACTTCAAA 515

Qy        490 GGTCCTGCGGTGTCGGTGGACACGGCTTGTTCCTCGTCGTTGGTGGCTTTGCATCTGGCC 549
              || ||      |  |  ||||||| || || ||||     ||| ||||  | ||    |||
Db        516 GGACCAAGCATTGCCCTGGACACAGCCTGCTCCTCCAGCTTGCTGGCACTACAGAATGCC 575

Qy        550 TGTCAGGCGTTGCGTGCGGGTGAGTGCTCGATGGCGCTTGCCGGGGGGTGTGACGGTGATG 609
              |  |||||  | |||   || || ||| |   ||| ||||  || ||  | ||  || |
Db        576 TACCAGGCCATCCGTAGTGGGGAATGCCCCGCGGCCCTTGTGGGTGGGATCACCCTGCTC 635

Qy        610 TCGTCTCCGGGTGCCTTCGTGGAGTTTTCGCGGCAGCGGGGTCTGGCCGCGGACG 664
              |   |||    |||  ||| ||||   |  ||   |    ||   | |||||
Db        636 CTGAAGCCGAACACCTCTGTGCAGTTCATGAAGCTTGGCATGCTCAGCTAGGACG 690
```

RESULT 7
AI390803
LOCUS AI390803 579 bp mRNA linear EST 15-MAR-2000
DEFINITION mb23h09.y1 Soares mouse p3NMF19.5 Mus musculus cDNA clone

```
          IMAGE:330305 5' similar to SW:FAS_RAT P12785 FATTY ACID SYNTHASE ;,
          mRNA sequence.
ACCESSION AI390803
VERSION   AI390803.1  GI:4216810
KEYWORDS  EST.
SOURCE    house mouse.
  ORGANISM  Mus musculus
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1  (bases 1 to 579)
  AUTHORS   Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T.,
          Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person
          ,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter
          ,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R.,
          Waterston,R. and Wilson,R.
  TITLE     The WashU-NCI Mouse EST Project 1999
  JOURNAL   Unpublished (1999)
COMMENT   Contact: Marra M/WashU-NCI Mouse EST Project 1999
          Washington University School of Medicine
          4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
          Tel: 314 286 1800
          Fax: 314 286 1810
          Email: mouseest@watson.wustl.edu
          This clone is available royalty-free through LLNL ; contact the
          IMAGE Consortium (info@image.llnl.gov) for further information.
          This read is a RESEQUENCE of a previously sequenced mouse clone
          This read has been verified (found to hit its original self in the
          correct orientation)
          MGI:211705
          Seq primer: -40RP from Gibco
          High quality sequence stop: 464
          POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .579
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:330305"
                     /clone_lib="Soares mouse p3NMF19.5"
                     /dev_stage="19.5 dpc total fetus"
                     /lab_host="DH10B (ampicillin resistant)"
                     /note="Vector: pT7T3D (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                     was primed with a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCATTTTTTTTTTTTTTTTTT 3'],
                     double-stranded cDNA was size selected, ligated to Eco RI
                     adapters (Pharmacia), digested with Not I and cloned into
                     the Not I and Eco RI sites of a modified pT7T3 vector
                     (Pharmacia). Library went through one round of
                     normalization to a Cot = 5. Library constructed by Bento
                     Soares and M.Fatima Bonaldo. RNA was kindly provided by
                     Dr. Minoru Ko (Wayne State University)."
BASE COUNT      116 a    166 c    165 g    132 t
ORIGIN
```

```
  Query Match             8.0%;  Score 103.4;  DB 9;  Length 579;
  Best Local Similarity   50.5%;  Pred. No. 3.8e-10;
  Matches  277;  Conservative    0;  Mismatches  271;  Indels    1;  Gaps    1;

Qy        210 CCACTTCGACGCCGAATTCTTCGGCATCAGCCCCCGCGAAGCCCTCGCCATGGACCCCCA 269
              | | |||||||||    || || ||  ||  ||||    | || | | | ||||||||||
Db          9 CAAGTTCGACGCCTCCTTTTTTGGGGTCCACCCCAAGCAGGCACACACAATGGACCCCCA 68

Qy        270 GCAACGACTCCTCCTCGAAACCGCCTGGGAAACCATCGAACACGCCGGCATCAACCCCCA 329
              ||  || || ||  | ||| | ||  ||| | || |   | |  || ||||||||
Db         69 GCTTCGGCTGCTGTTGGAAGTCAGCTATGAAGCAATTGTGGATGGAGGTATCAACCCAGC 128

Qy        330 CACCCTCCACGGCACCCCCACCGGAGTCTTCACCGGCACCAACGGACAGGACTACGCACT 389
              | | ||||  || ||   ||| || ||||     ||    |  ||    ||     |
Db        129 CTCACTCCGAGGAACGAACACTGGCGTCTGGGTGGGTGTGAGTGGTTCAGAGGCATCCGA 188

Qy        390 TCGCGTGCACAACGCGGGCCAGTCAACCGATGGTTTCGCACTGACCGGAACCGCCGGCAG 449
                | |   ||  |   | || |     || | |   ||   |   ||  | |
Db        189 GGCCCTTAGCAGAGATCCCGAGACGCTTCTGGGCTACAGCATGGTGGCTGCCAGC-GTGC 247

Qy        450 CGTCATCTCCGGTCGTATCTCGTACACGTTTGGTTTTGAGGGTCCTGCGGTGTCGGTGGA 509
              | ||  ||   ||  |||| | |   || |  ||  | || ||     |  |  ||||
Db        248 AATGATGGCCAACCGGCTCTCTTTCTTCTTCGACTTCAAAGGACCAAGCATTGCCCTGGA 307

Qy        510 CACGGCTTGTTCCTCGTCGTTGGTGGCTTTGCATCTGGCCTGTCAGGCGTTGCGTGCGGG 569
              ||| || || |||||    ||| ||||  | ||    ||||  |||||  | |||   ||
Db        308 CACAGCCTGCTCCTCCAGCTTGCTGGCACTACAGAATGCCTACCAGGCCATCCGTAGTGG 367

Qy        570 TGAGTGCTCGATGGCGCTTGCCGGGGGTGTGACGGTGATGTCGTCTCCGGGTGCCTTCGT 629
              || ||| |   ||| |||| || ||  | |   || |   |   |||   |||  ||
Db        368 GGAATGCCCCGCGGCCCTTGTGGGTGGGATCAACCTGCTCCTGAAGCCGAACACCTCTGT 427

Qy        630 GGAGTTTTCGCGGCAGCGGGGTCTGGCCGCGGACGGGCATTGCAAGGCGTTCTCGGCGGC 689
              | ||||   |  || | |   ||   | |||||||    ||||   | ||   |   |
Db        428 GCAGTTCATGAAGCTGGGCATGCTCAGCCCGGACGGCACCTGCAGATCCTTTGATGATTC 487

Qy        690 GGCGGACGGGACCGGCTGGGGTGAGGGTGTGGGGATGCTGCTGGTGGAGCGGCTCTCCGA 749
              | |   ||     |  |  ||||| ||| |      | ||| ||     |   |||
Db        488 AGGGAGTGGATATTGTCGCTCTGAGGCTGTTGTAGCAGTTCTGCTGACTAAGAAGTCCCT 547

Qy        750 CGCCCATCG 758
              || |  ||
Db        548 GGCTCGGCG 556
```

```
RESULT 8
BB866094
LOCUS       BB866094                640 bp    mRNA    linear   EST 27-NOV-2001
DEFINITION  BB866094 RIKEN full-length enriched, CRL-2116 JC cDNA Mus musculus
          cDNA clone G431004B12 5', mRNA sequence.
ACCESSION   BB866094
VERSION     BB866094.1  GI:17112304
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 640)
  AUTHORS   Akimura,T., Arakawa,T., Carninci,P., Furuno,M., Hanagaki,T.,
          Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T., Imotani,K., Ishii
          ,Y., Ito,M., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Matsuyama,T.,
          Nakamura,M., Nishi,K., Nomura,K., Numasaki,R., Okazaki,Y., Okido,T.
          , Saito,R., Sakai,C., Sakai,K., Sakazume,N., Sasaki,D., Sato,K.,
          Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagawa
          ,A., Takahashi,F., Takaku-Akahira,S., Tanaka,T., Tomaru,A., Toya,T.
          , Watahiki,A., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y.
  TITLE     RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura,T., et al.
          2001)
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Yoshihide Hayashizaki
          Laboratory for Genome Exploration Research Group, RIKEN Genomic
          Sciences Center(GSC), Yokohama Institute
          The Institute of Physical and Chemical Research (RIKEN)
          1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
          Tel: 81-45-503-9222
          Fax: 81-45-503-9216
          Email: genome-res@gsc.riken.go.jp,
          URL:http://genome.gsc.riken.go.jp/
          Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh
          ,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
           Normalization and subtraction of cap-trapper-selected cDNAs to
          prepare full-length cDNA libraries for rapid discovery of new
          genes. Genome Res. . 10 (10), 1617-1630 (2000)
           wagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
          Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura
          ,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and
          Hayashizaki,Y.
           RIKEN integrated sequence analysis (RISA) system--384-format
          sequencing pipeline with 384 multicapillary sequencer. Genome Res. .
          10 (11), 1757-1771 (2000)
           Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara
          ,Y. and Hayashizaki,Y.
```

```
            Computer-based methods for the mouse full-length cDNA
           encyclopedia: real-time sequence clustering for construction of a
           nonredundant cDNA library. Genome Res. . 11 (2), 281-289 (2001)
            Please visit our web site (http://genome.gsc.riken.go.jp) for
           further details.
            e mouse tissues.
FEATURES             Location/Qualifiers
     source          1. .640
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="G431004B12"
                     /clone_lib="RIKEN full-length enriched, CRL-2116 JC cDNA"
                     /cell_line="CRL-2116 JC"
                     /note="pooled cell lines ; (cell_line=CRL-1751 WEHI 164),
                     (cell_line=CRL-2116 JC), (cell_line=RCB-0035 WEHI-3),
                     (cell_line=RCB-0464 Meth-A), (cell_line=RCB-0545 OHTA),
                     (cell_line=RCB-0559 K-1 . F1), (cell_line=RCB-1283 B16
                     melanoma), (cell_type=B cells, cell_line=CRL-1702 WEHI 231
                     ), (cell_type=Leydig cells, cell_line=CRL-2065 MLTC-1),
                     (cell_type=Nullipotent stem cell, cell_line=CRL-2070 NE),
                     (tissue_type=bladder, cell_line=RCB-0544 MBT-2),
                     (tissue_type=bone marrow, cell_type=stroma cell,
                     cell_line= CRL-2028 SR-4987), (tissue_type=colon,
                     cell_line=RCB-0549 Cle-H3), (tissue_type=kidney,
                     cell_line=CCL-142 RAG), (tissue_type=submandibular gland,
                     cell_line=CRL-1734 SCA-9 clone 15), (strain=BALB/C,
                     cell_type=B cells, cell_line=CRL-1669 BCL1 Clone 13.20-3B3
                     ), (strain=C3H, tissue_type=brain, cell_line=CRL-1443
                     BC3H1)"
BASE COUNT      138 a    172 c    195 g    134 t      1 others
ORIGIN

  Query Match             7.8%;  Score 101;  DB 10;  Length 640;
  Best Local Similarity   51.6%;  Pred. No. 1.1e-09;
  Matches  230;  Conservative    0;  Mismatches  216;  Indels    0;  Gaps    0;

Qy        190 GGATTCCTCTACGACGCAGGCCACTTCGACGCCGAATTCTTCGGCATCAGCCCCCGCGAA 249
              |||   ||  | ||       | | |||||||||   || || ||  ||  ||||    |
Db        195 GGAAAGCTGAAGGATCTCTCCAAGTTCGACGCCTCCTTTTTTGGGGTCCACCCCAAGCAG 254

Qy        250 GCCCTCGCCATGGACCCCCAGCAACGACTCCTCCTCGAAACCGCCTGGGAAACCATCGAA 309
              || | | | |||||||||||||  || || ||  | |||  |  ||  ||| | || |
Db        255 GCACACACAATGGACCCCCAGCTTCGGCTGCTGTTGGAAGTCAGCTATGAAGCAATTGTG 314

Qy        310 CACGCCGGCATCAACCCCCACACCCTCCACGGCACCCCCACCGGAGTCTTCACCGGCACC 369
              | |  || |||||||  | | ||||  || ||   ||| || ||||      ||
Db        315 GATGGAGGTATCAACCCAGCCTCACTCCGAGGAACGAACACTGGCGTCTGGGTGGGTGTG 374

Qy        370 AACGGACAGGACTACGCACTTCGCGTGCACAACGCGGGCCAGTCAACCGATGGTTTCGCA 429
              |  ||   ||     |     | |   ||  |    | || |          || | |
Db        375 AGTGGTTCAGAGGCATCCGAGGCCCTTAGCAGAGATCCCGAGACGCTTCTGGGCTACAGC 434

Qy        430 CTGACCGGAACCGCCGGCAGCGTCATCTCCGGTCGTATCTCGTACACGTTTGGTTTTGAG 489
              ||   ||   |    |      | ||  ||   ||  |||| | |   || |  ||  |
Db        435 ATGGTGGGCTGCCAGCGTGCAATGATGGCCAACCGGCTCTCTTTCTTCTTCGACTTCAAA 494

Qy        490 GGTCCTGCGGTGTCGGTGGACACGGCTTGTTCCTCGTCGTTGGTGGCTTTGCATCTGGCC 549
              || ||     |  |  ||||||| || || |||||     ||| ||||  | ||    |||
Db        495 GGACCAAGCATTGCCCTGGACACAGCCTGCTCCTCCAGCTTGCTGGCACTACAGAATGCC 554

Qy        550 TGTCAGGCGTTGCGTGCGGGTGAGTGCTCGATGGCGCTTGCCGGGGGTGTGACGGTGATG 609
              |  |||||  | |||   || || ||| |   ||| |||| || ||  | |   || |
Db        555 TACCAGGCCATCCGTAGTGGGGAATGCNCCGCGGCCCTTGTGGGTGGGATCAACCTGCTC 614

Qy        610 TCGTCTCCGGGTGCCTTCGTGGAGTT 635
                |   |||    |||  ||| ||||
Db        615 CTGAAGCCGAACACCTCTGTGCAGTT 640


RESULT 9
BB866070
LOCUS       BB866070                682 bp    mRNA    linear   EST 27-NOV-2001
DEFINITION  BB866070 RIKEN full-length enriched, submandibular gland CRL-1734
            SCA-9 clone 15 cDNA Mus musculus cDNA clone G431003P17 5', mRNA
            sequence.
ACCESSION   BB866070
VERSION     BB866070.1  GI:17112280
KEYWORDS    EST.
SOURCE      house mouse..
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 682)
  AUTHORS   Akimura,T., Arakawa,T., Carninci,P., Furuno,M., Hanagaki,T.,
            Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T., Imotani,K., Ishii
            ,Y., Ito,M., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Matsuyama,T.,
            Nakamura,M., Nishi,K., Nomura,K., Numasaki,R., Okazaki,Y., Okido,T.
            , Saito,R., Sakai,C., Sakai,K., Sakazume,N., Sasaki,D., Sato,K.,
            Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagawa
            ,A., Takahashi,F., Takaku-Akahira,S., Tanaka,T., Tomaru,A., Toya,T.
            , Watahiki,A., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y.
  TITLE     RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura,T., et al.
            2001)
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/
            Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh
            ,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
             Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new
            genes. Genome Res. . 10 (10), 1617-1630 (2000)
             wagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
            Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura
            ,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and
            Hayashizaki,Y.
             RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer. Genome Res.
            10 (11), 1757-1771 (2000)
             Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara
            ,Y. and Hayashizaki,Y.
             Computer-based methods for the mouse full-length cDNA
            encyclopedia: real-time sequence clustering for construction of a
            nonredundant cDNA library. Genome Res. . 11 (2), 281-289 (2001)
             Please visit our web site (http://genome.gsc.riken.go.jp) for
            further details.
             e mouse tissues.
FEATURES             Location/Qualifiers
     source          1. .682
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="G431003P17"
                     /clone_lib="RIKEN full-length enriched, submandibular
                     gland CRL-1734 SCA-9 clone 15 cDNA"
                     /tissue_type="submandibular gland"
                     /cell_line="CRL-1734 SCA-9 clone 15"
                     /note="pooled cell lines ; (cell_line=CRL-1751 WEHI 164),
                     (cell_line=CRL-2116 JC), (cell_line=RCB-0035 WEHI-3),
                     (cell_line=RCB-0464 Meth-A), (cell_line=RCB-0545 OHTA),
                     (cell_line=RCB-0559 K-1 . F1), (cell_line=RCB-1283 B16
                     melanoma), (cell_type=B cells, cell_line=CRL-1702 WEHI 231
                     ), (cell_type=Leydig cells, cell_line=CRL-2065 MLTC-1),
                     (cell_type=Nullipotent stem cell, cell_line=CRL-2070 NE),
                     (tissue_type=bladder, cell_line=RCB-0544 MBT-2),
                     (tissue_type=bone marrow, cell_type=stroma cell,
                     cell_line= CRL-2028 SR-4987), (tissue_type=colon,
                     cell_line=RCB-0549 Cle-H3), (tissue_type=kidney,
                     cell_line=CCL-142 RAG), (tissue_type=submandibular gland,
                     cell_line=CRL-1734 SCA-9 clone 15), (strain=BALB/C,
                     cell_type=B cells, cell_line=CRL-1669 BCL1 Clone 13.20-3B3
```

```
                ), (strain=C3H, tissue_type=brain, cell_line=CRL-1443
                BC3H1)"
BASE COUNT      138 a    192 c    209 g    141 t      2 others
ORIGIN

  Query Match             7.8%;  Score 100;  DB 10;  Length 682;
  Best Local Similarity   51.3%;  Pred. No. 1.8e-09;
  Matches  229;  Conservative    0;  Mismatches  217;  Indels    0;  Gaps    0;

Qy        190 GGATTCCTCTACGACGCAGGCCACTTCGACGCCGAATTCTTCGGCATCAGCCCCCGCGAA 249
              |||    ||  | ||       | | |||||||||   || || ||  ||  ||||    |
Db        237 GGAAAGCTGAAGGATCTCTCCAAGTTCGACGCCTCCTTTTTTTGGGGTCCACCCCAAGCAG 296

Qy        250 GCCCTCGCCATGGACCCCCAGCAACGACTCCTCCTCGAAACCGCCTGGGAAACCATCGAA 309
              || | | | ||||||||||||||  || || ||  | |||  |  ||  ||| | || |
Db        297 GCACACACAATGGACCCCCAGCTTCGGCTGCTGTTGGAAGTCAGCTATGAAGCAATTGTG 356

Qy        310 CACGCCGGCATCAACCCCCACACCCTCCACGGCACCCCCACCGGAGTCTTCACCGGCACC 369
              | |  || ||||||||   | | ||||  || ||    ||| || ||||      ||
Db        357 GATGGAGGTATCAACCCAGCCTCACTCCGAGGAACGAACACTGGCGTCTGGGTGGGTGTG 416

Qy        370 AACGGACAGGACTACGCACTTCGCGTGCACAACGCGGGCCAGTCAACCGATGGTTTCGCA 429
              |  ||    ||      |     | |   ||  |     | || |         || | |
Db        417 AGTGGTTCAGAGGCATCCGAGGCCCTTAGCAGAGATCCCGAGACGCTTCTGGGCTACAGC 476

Qy        430 CTGACCGGAACCGCCGGCAGCGTCATCTCCGGTCGTATCTCGTACACGTTTGGTTTTGAG 489
              ·||   ||   |   |   | | ||   ||   ||  |||| | |   || |  ||  |
Db        477 ATGGTGGGCTGCCAGCGTGCCATGATGGCCAACCGGCTCTCTTTCTTCTTCGACTTCAAA 536

Qy        490 GGTCCTGCGGTGTCGGTGGACACGGCTTGTTCCTCGTCGTTGGTGGCTTTGCATCTGGCC 549
              || ||     |  |  ||||||| || || |||||    ||| ||||  | ||    |||
Db        537 GGACCAAGCATTGCCCTGGACACAGCCTGCTCCTCCAGCTTGCTGGCNCTACAGAATGCC 596

Qy        550 TGTCAGGCGTTGCGTGCGGGTGAGTGCTCGATGGCGCTTGCCGGGGGTGTGACGGTGATG 609
              |  |||||  | ||     || || |||  |   ||| ||||  || ||  | |   || |
Db        597 TACCAGGCCATCCGGAGTGGGGAATGCCCCGCGGCCCTTGTGGGTGGGATCAACCTGCTN 656

Qy        610 TCGTCTCCGGGTGCCTTCGTGGAGTT 635
                  |||   |||  ||| ||||
Db        657 CTTGAGCCGAACACCTCTGTGCAGTT 682


RESULT 10
AW659924
LOCUS       AW659924                546 bp    mRNA    linear   EST 25-APR-2001
DEFINITION  97589 MARC 1BOV Bos taurus cDNA 5', mRNA sequence.
ACCESSION   AW659924
VERSION     AW659924.1  GI:7425688
KEYWORDS    EST.
SOURCE      cow.
  ORGANISM  Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
            Bovidae; Bovinae; Bos.
REFERENCE   1  (bases 1 to 546)
  AUTHORS   Smith,T.P.L., Grosse,W.M., Freking,B.A., Roberts,A.J., Stone,R.T.,
            Casas,E., Wray,J.E., White,J., Cho,J., Fahrenkrug,S.C., Bennett
            ,G.L., Heaton,M.P., Laegreid,W.W., Rohrer,G.A., Chitko-McKown,C.G.,
            Pertea,G., Holt,I., Karamycheva,S., Liang,F., Quackenbush,J. and
            Keele,J.W.
  TITLE     Sequence evaluation of four pooled-tissue normalized bovine cDNA
            libraries and construction of a gene index for cattle
  JOURNAL   Genome Res. 11 (4), 626-630 (2001)
  MEDLINE   21180013
COMMENT     Contact: Smith TPL
            USDA, ARS, US Meat Animal Research Center
            PO Box 166, Clay Center, NE 68933-0166, USA
            Tel: 402 762 4366
            Fax: 402 762 4390
            Email: smith@email.marc.usda.gov
            Single pass sequencing. Bases called and alt_trimmed with phred
            v0.980904.e. Vector identified by cross_match with the -minscore 18
            and -minmatch 12 options.
            PCR PRimers
            FORWARD: AGGAAACAGCTATGACCAT
            BACKWARD: GTTTTCCCAGTCACGACG
            Plate: 89  row: H  column: 2
            Seq primer: ATTTAGGTGACACTATAG.
FEATURES              Location/Qualifiers
     source           1. .546
                      /organism="Bos taurus"
                      /db_xref="taxon:9913"
                      /clone_lib="MARC 1BOV"
                      /tissue_type="pooled"
                      /lab_host="DH10B"
                      /note="Vector: pCMV SPORT6; Site_1: NotI; Site_2: SalI;
                      Library made from pooled tissue from lymph node, ovary,
                      fat, hypothalamus, and pituitary."
BASE COUNT       99 a    159 c    177 g    111 t
ORIGIN

  Query Match             7.7%;  Score 99.4;  DB 10;  Length 546;
  Best Local Similarity   53.1%;  Pred. No. 2.2e-09;
  Matches  211;  Conservative    0;  Mismatches  186;  Indels    0;  Gaps    0;

Qy        196 CTCTACGACGCAGGCCACTTCGACGCCGAATTCTTCGGCATCAGCCCCCGCGAAGCCCTC 255
              ||  | |||     ||  || |||||    ||||||||  ||  ||||    | || | |
Db        150 CTGAAGGACCTGTCCCGGTTTGACGCTTCCTTCTTCGGGGTCCACCCCAAGCAGGCACAC 209

Qy        256 GCCATGGACCCCCAGCAACGACTCCTCCTCGAAACCGCCTGGGAAACCATCGAACACGCC 315
                |||||||||||||  ||  | || || ||   | |||  ||  |||| |   | ||
Db        210 AATATGGACCCCCAGCTCCGCTTGCTGCTGGAGGTCACCTACGAGGCCATTGTGGATGCA 269

Qy        316 GGCATCAACCCCCACACCCTCCACGGCACCCCCACCGGAGTCTTCACCGGCACCAACGGA 375
              ||||||||||     || | |  || ||   |||||| ||||      ||    |  ||
Db        270 GGCATCAACCCAGCTTCCATTCGGGGGACGAACACCGGTGTCTGGGTGGGTGTGAGTGGC 329

Qy        376 CAGGACTACGCACTTCGCGTGCACAACGCGGGCCAGTCAACCGATGGTTTCGCACTGACC 435
                ||     ||     ||  |   |     || |   ||  || | |.   ||
Db        330 TCTGAGGCTTCAGAGGCTCTGAGCCGAGACCCTGAGACCCTCGTGGGCTACAGCATGGTG 389

Qy        436 GGAACCGCCGGCAGCGTCATCTCCGGTCGTATCTCGTACACGTTTGGTTTTGAGGGTCCT 495
              ||   |   |  | |  |  ||   ||  |||| | |   ||||  ||  | || ||
Db        390 GGCTGCCAGCGTGCCATGTTGGCCAACCGCCTCTCCTTCTTCTTTGACTTCAAAGGGCCC 449

Qy        496 GCGGTGTCGGTGGACACGGCTTGTTCCTCGTCGTTGGTGGCTTTGCATCTGGCCTGTCAG 555
              |  |  |||||||||| || |||||     || |||| ||||  ||||  |||
Db        450 AGCATCACCCTGGACACGGCATGCTCCTCCAGCCTGCTGGCCCTGCAGAGGGCCTACCAG 509

Qy        556 GCGTTGCGTGCGGGTGAGTGCTCGATGGCGCTTGCCG 592
              ||  | |    || |||||| | |||||  ||| ||
Db        510 GCCATCCAGAGAGGGGAGTGCGCCATGGCCATTGTCG 546


RESULT 11
BB866199
LOCUS       BB866199                638 bp    mRNA    linear   EST 27-NOV-2001
DEFINITION  BB866199 RIKEN full-length enriched, CRL-2116 JC cDNA Mus musculus
            cDNA clone G431004J01 5', mRNA sequence.
ACCESSION   BB866199
VERSION     BB866199.1  GI:17112409
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 638)
  AUTHORS   Akimura,T., Arakawa,T., Carninci,P., Furuno,M., Hanagaki,T.,
            Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T., Imotani,K., Ishii
            ,Y., Ito,M., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Matsuyama,T.,
            Nakamura,M., Nishi,K., Nomura,K., Numasaki,R., Okazaki,Y., Okido,T.
            , Saito,R., Sakai,C., Sakai,K., Sakazume,N., Sasaki,D., Sato,K.,
            Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagawa
            ,A., Takahashi,F., Takaku-Akahira,S., Tanaka,T., Tomaru,A., Toya,T.
            , Watahiki,A., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y.
  TITLE     RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura,T., et al.
```

```
            2001)
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/
            Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh
            ,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
             Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new
            genes. Genome Res. . 10 (10), 1617-1630 (2000)
             wagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
            Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura
            ,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and
            Hayashizaki,Y.
             RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer. Genome Res. .
            10 (11), 1757-1771 (2000)
             Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara
            ,Y. and Hayashizaki,Y.
             Computer-based methods for the mouse full-length cDNA
            encyclopedia: real-time sequence clustering for construction of a
            nonredundant cDNA library. Genome Res. . 11 (2), 281-289 (2001)
             Please visit our web site (http://genome.gsc.riken.go.jp) for
            further details.
             e mouse tissues.
FEATURES             Location/Qualifiers
     source          1. .638
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="G431004J01"
                     /clone_lib="RIKEN full-length enriched, CRL-2116 JC cDNA"
                     /cell_line="CRL-2116 JC"
                     /note="pooled cell lines ; (cell_line=CRL-1751 WEHI 164),
                     (cell_line=CRL-2116 JC), (cell_line=RCB-0035 WEHI-3),
                     (cell_line=RCB-0464 Meth-A), (cell_line=RCB-0545 OHTA),
                     (cell_line=RCB-0559 K-1 . F1), (cell_line=RCB-1283 B16
                     melanoma), (cell_type=B cells, cell_line=CRL-1702 WEHI 231
                     ), (cell_type=Leydig cells, cell_line=CRL-2065 MLTC-1),
                     (cell_type=Nullipotent stem cell, cell_line=CRL-2070 NE),
                     (tissue_type=bladder, cell_line=RCB-0544 MBT-2),
                     (tissue_type=bone marrow, cell_type=stroma cell,
                     cell_line= CRL-2028 SR-4987), (tissue_type=colon,
                     cell_line=RCB-0549 Cle-H3), (tissue_type=kidney,
                     cell_line=CCL-142 RAG), (tissue_type=submandibular gland,
                     cell_line=CRL-1734 SCA-9 clone 15), (strain=BALB/C,
                     cell_type=B cells, cell_line=CRL-1669 BCL1 Clone 13.20-3B3
                     ), (strain=C3H, tissue_type=brain, cell_line=CRL-1443
                     BC3H1)"
BASE COUNT      136 a    173 c    195 g    134 t
ORIGIN

  Query Match             7.6%;  Score 98.4;  DB 10;  Length 638;
  Best Local Similarity   51.4%;  Pred. No. 3.5e-09;
  Matches  228;  Conservative    0;  Mismatches  216;  Indels    0;  Gaps    0;

Qy        190 GGATTCCTCTACGACGCAGGCCACTTCGACGCCGAATTCTTCGGCATCAGCCCCCGCGAA 249
              |||   ||  | ||       | | |||||||||   || || ||  ||  ||||    |
Db        195 GGAAAGCTGAAGGATCTCTCCAAGTTCGACGCCTCCTTTTTTGGGGTCCACCCCAAGCAG 254

Qy        250 GCCCTCGCCATGGACCCCCAGCAACGACTCCTCCTCGAAACCGCCTGGGAAACCATCGAA 309
              || | | | |||||||||||||  || || ||  | |||  |  ||  ||| | || |
Db        255 GCACACACAATGGACCCCCAGCTTCGGCTGCTGTTGGAAGTCAGCTATGAAGCAATTGTG 314

Qy        310 CACGCCGGCATCAACCCCCACACCCTCCACGGCACCCCCACCGGAGTCTTCACCGGCACC 369
              | |  || ||||||||   | | ||||  || ||   ||| || ||||     ||
Db        315 GATGGAGGTATCAACCCAGCCTCACTCCGAGGAACGAACACTGGCGTCTGGGTGGGTGTG 374

Qy        370 AACGGACAGGACTACGCACTTCGCGTGCACAACGCGGGCCAGTCAACCGATGGTTTCGCA 429
              | ||   ||      |     | |  ||  |    | || |        || | |
Db        375 AGTGGTTCAGAGGCATCCGAGGCCCTTAGCAGAGATCCCGAGACGCTTCTGGGCTACAGC 434

Qy        430 CTGACCGGAACCGCCGGCAGCGTCATCTCCGGTCGTATCTCGTACACGTTTGGTTTTGAG 489
              ||  ||  |   |      | ||  ||  ||  |||| | |  || | ||  |
Db        435 ATGGTGGGCTGCCAGCGTGCAATGATGGCCAACCGGCTCTCTTTCTTCTTCGACTTCAAA 494

Qy        490 GGTCCTGCGGTGTCGGTGGACACGGCTTGTTCCTCGTCGTTGGTGGCTTTGCATCTGGCC 549
              || ||     |  | ||||||| || || |||||    ||| ||||  | ||   |||
Db        495 GGACCAAGCATTGCCCTGGACACAGCCTGCTCCTCCAGCTTGCTGGCACTACAGAATGCC 554

Qy        550 TGTCAGGCGTTGCGTGCGGGTGAGTGCTCGATGGCGCTTGCCGGGGGTGTGACGGTGATG 609
              |  |||||  | |||   || || ||| |   ||| ||||  || ||  | |   || |
Db        555 TACCAGGCCATCCGTAGTGGGGAATGCCCCGCGGCCCTTGTGGGTGGGATCAACCTGCTC 614

Qy        610 TCGTCTCCGGGTGCCTTCGTGGAG 633
              |   |||    |||  ||| ||
Db        615 CTGAAGCCGATCACCTCTGTGCAG 638


RESULT 12
BB865388
LOCUS       BB865388                681 bp    mRNA    linear   EST 27-NOV-2001
DEFINITION  BB865388 RIKEN full-length enriched, pooled cell lines Mus musculus
            cDNA clone G431001F02 5', mRNA sequence.
ACCESSION   BB865388
VERSION     BB865388.1  GI:17111598
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 681)
  AUTHORS   Akimura,T., Arakawa,T., Carninci,P., Furuno,M., Hanagaki,T.,
            Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T., Imotani,K., Ishii
            ,Y., Ito,M., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Matsuyama,T.,
            Nakamura,M., Nishi,K., Nomura,K., Numasaki,R., Okazaki,Y., Okido,T.
            , Saito,R., Sakai,C., Sakai,K., Sakazume,N., Sasaki,D., Sato,K.,
            Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagawa
            ,A., Takahashi,F., Takaku-Akahira,S., Tanaka,T., Tomaru,A., Toya,T.
            , Watahiki,A., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y.
  TITLE     RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura,T., et al.
            2001)
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/
            Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh
            ,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
             Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new
            genes. Genome Res. . 10 (10), 1617-1630 (2000)
             wagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
            Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura
            ,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and
            Hayashizaki,Y.
             RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer. Genome Res. .
            10 (11), 1757-1771 (2000)
             Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara
            ,Y. and Hayashizaki,Y.
             Computer-based methods for the mouse full-length cDNA
            encyclopedia: real-time sequence clustering for construction of a
            nonredundant cDNA library. Genome Res. . 11 (2), 281-289 (2001)
             Please visit our web site (http://genome.gsc.riken.go.jp) for
            further details.
```

```
          e mouse tissues.
FEATURES             Location/Qualifiers
     source          1. .681
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="G431001F02"
                     /clone_lib="RIKEN full-length enriched, pooled cell lines"
                     /note="pooled cell lines ; (cell_line=CRL-1751 WEHI 164),
                     (cell_line=CRL-2116 JC), (cell_line=RCB-0035 WEHI-3),
                     (cell_line=RCB-0464 Meth-A), (cell_line=RCB-0545 OHTA),
                     (cell_line=RCB-0559 K-1 . F1), (cell_line=RCB-1283 B16
                     melanoma), (cell_type=B cells, cell_line=CRL-1702 WEHI 231
                     ), (cell_type=Leydig cells, cell_line=CRL-2065 MLTC-1),
                     (cell_type=Nullipotent stem cell, cell_line=CRL-2070 NE),
                     (tissue_type=bladder, cell_line=RCB-0544 MBT-2),
                     (tissue_type=bone marrow, cell_type=stroma cell,
                     cell_line= CRL-2028 SR-4987), (tissue_type=colon,
                     cell_line=RCB-0549 Cle-H3), (tissue_type=kidney,
                     cell_line=CCL-142 RAG), (tissue_type=submandibular gland,
                     cell_line=CRL-1734 SCA-9 clone 15), (strain=BALB/C,
                     cell_type=B cells, cell_line=CRL-1669 BCL1 Clone 13.20-3B3
                     ), (strain=C3H, tissue_type=brain, cell_line=CRL-1443
                     BC3H1)"
BASE COUNT      141 a    192 c    209 g    139 t
ORIGIN

  Query Match             7.4%;  Score 96;  DB 10;  Length 681;
  Best Local Similarity   51.1%;  Pred. No. 1e-08;
  Matches  225;  Conservative    0;  Mismatches  215;  Indels    0;  Gaps    0;

Qy        190 GGATTCCTCTACGACGCAGGCCACTTCGACGCCGAATTCTTCGGCATCAGCCCCCGCGAA 249
              |||   ||  | ||       | | |||||||||   || || ||  ||  ||||    |
Db        242 GGAAAGCTGAAGGATCTCTCCAAGTTCGACGCCTCCTTTTTTGGGGTCCACCCCAAGCAG 301

Qy        250 GCCCTCGCCATGGACCCCCAGCAACGACTCCTCCTCGAAACCGCCTGGGAAACCATCGAA 309
              || | | | ||||||||||||||  || || ||  | |||  |  ||  ||| | || |
Db        302 GCACACACAATGGACCCCCAGCTTCGGCTGCTGTTGGAAGTCAGCTATGAAGCAATTGTG 361

Qy        310 CACGCCGGCATCAACCCCCACACCCTCCACGGCACCCCCACCGGAGTCTTCACCGGCACC 369
               | |  || ||||||||   | | ||||  || ||   ||| || ||||     ||
Db        362 GATGGAGGTATCAACCCAGCCTCACTCCGAGGAACGAACACTGGCGTCTGGGTGGGTGTG 421

Qy        370 AACGGACAGGACTACGCACTTCGCGTGCACAACGCGGGCCAGTCAACCGATGGTTTCGCA 429
              | ||   ||       |     | |   ||  |    | || |        || | |
Db        422 AGTGGTTCAGAGGCATCCGAGGCCCTTAGCAGAGATCCCGAGACGCTTCTGGGCTACAGC 481

Qy        430 CTGACCGGAACCGCCGGCAGCGTCATCTCCGGTCGTATCTCGTACACGTTTGGTTTTGAG 489
               ||   ||  |    |     | ||  ||   ||  |||| | |   || |  ||  |
Db        482 ATGGTGGGCTGCCAGCGTGCAATGATGGCCAACCGGCTCTCTTTCTTCTTCGACTTCAAA 541

Qy        490 GGTCCTGCGGTGTCGGTGGACACGGCTTGTTCCTCGTCGTTGGTGGCTTTGCATCTGGCC 549
              || ||     |  |  ||||||| || || |||||    ||| ||||  | ||    |||
Db        542 GGACCAAGCATTGCCCTGGACACAGCCTGCTCCTCCAGCTTGCTGGCACTACAGAATGCC 601

Qy        550 TGTCAGGCGTTGCGTGCGGGTGAGTGCTCGATGGCGCTTGCCGGGGGGTGTGACGGTGATG 609
              |  |||||  | |||   || || ||| |   ||| ||||  || ||  | |   || |
Db        602 TACCAGGCCATCCGTAGTGGGGAATGCCCCGCGGCCCTTGTGGGTGGGATCAACCTGCTC 661

Qy        610 TCGTCTCCGGGTGCCTTCGT 629
               |    |||    |||  ||
Db        662 CCTGAGCCGAACACCTCTGT 681


RESULT 13
BB865339
LOCUS       BB865339                 701 bp    mRNA    linear   EST 27-NOV-2001
DEFINITION  BB865339 RIKEN full-length enriched, RCB-0545 OHTA cDNA Mus
            musculus cDNA clone G431001C12 5', mRNA sequence.
ACCESSION   BB865339
VERSION     BB865339.1  GI:17111549
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 701)
  AUTHORS   Akimura,T., Arakawa,T., Carninci,P., Furuno,M., Hanagaki,T.,
            Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T., Imotani,K., Ishii
            ,Y., Ito,M., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Matsuyama,T.,
            Nakamura,M., Nishi,K., Nomura,K., Numasaki,R., Okazaki,Y., Okido,T.
            , Saito,R., Sakai,C., Sakai,K., Sakazume,N., Sasaki,D., Sato,K.,
            Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagawa
            ,A., Takahashi,F., Takaku-Akahira,S., Tanaka,T., Tomaru,A., Toya,T.
            , Watahiki,A., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y.
  TITLE     RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura,T., et al.
            2001)
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/
            Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh
            ,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
             Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new
            genes. Genome Res. . 10 (10), 1617-1630 (2000)
             wagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
            Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura
            ,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and
            Hayashizaki,Y.
             RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer. Genome Res.
            10 (11), 1757-1771 (2000)
             Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara
            ,Y. and Hayashizaki,Y.
             Computer-based methods for the mouse full-length cDNA
            encyclopedia: real-time sequence clustering for construction of a
            nonredundant cDNA library. Genome Res. . 11 (2), 281-289 (2001)
             Please visit our web site (http://genome.gsc.riken.go.jp) for
            further details.
             e mouse tissues.
FEATURES             Location/Qualifiers
     source          1. .701
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="G431001C12"
                     /clone_lib="RIKEN full-length enriched, RCB-0545 OHTA
                     cDNA"
                     /cell_line="RCB-0545 OHTA"
                     /note="pooled cell lines ; (cell_line=CRL-1751 WEHI 164),
                     (cell_line=CRL-2116 JC), (cell_line=RCB-0035 WEHI-3),
                     (cell_line=RCB-0464 Meth-A), (cell_line=RCB-0545 OHTA),
                     (cell_line=RCB-0559 K-1 . F1), (cell_line=RCB-1283 B16
                     melanoma), (cell_type=B cells, cell_line=CRL-1702 WEHI 231
                     ), (cell_type=Leydig cells, cell_line=CRL-2065 MLTC-1),
                     (cell_type=Nullipotent stem cell, cell_line=CRL-2070 NE),
                     (tissue_type=bladder, cell_line=RCB-0544 MBT-2),
                     (tissue_type=bone marrow, cell_type=stroma cell,
                     cell_line= CRL-2028 SR-4987), (tissue_type=colon,
                     cell_line=RCB-0549 Cle-H3), (tissue_type=kidney,
                     cell_line=CCL-142 RAG), (tissue_type=submandibular gland,
                     cell_line=CRL-1734 SCA-9 clone 15), (strain=BALB/C,
                     cell_type=B cells, cell_line=CRL-1669 BCL1 Clone 13.20-3B3
                     ), (strain=C3H, tissue_type=brain, cell_line=CRL-1443
                     BC3H1)"
BASE COUNT      145 a    195 c    216 g    142 t      3 others
ORIGIN

  Query Match             7.4%;  Score 95.8;  DB 10;  Length 701;
  Best Local Similarity   51.3%;  Pred. No. 1.1e-08;
  Matches  220;  Conservative    0;  Mismatches  209;  Indels    0;  Gaps    0;
```

```
Qy        190 GGATTCCTCTACGACGCAGGCCACTTCGACGCCGAATTCTTCGGCATCAGCCCCCGCGAA 249
              |||   ||  | ||       | | |||||||||   || || ||  ||  ||||    |
Db        242 GGAAAGCTGAAGGATCTCTCCAAGTTCGACGCCTCCTTTTTTGGGGTCCACCCCAAGCAG 301

Qy        250 GCCCTCGCCATGGACCCCCAGCAACGACTCCTCCTCGAAACCGCCTGGGAAACCATCGAA 309
              || | | ||||||||||||||  || || ||  | |||  |  ||  ||| | || |
Db        302 GCACACACAATGGACCCCCAGCTCCGGCTGCTGTTGGAAGTCAGCTATGAAGCAATTGTG 361

Qy        310 CACGCCGGCATCAACCCCCACACCCTCCACGGCACCCCCACCGGAGTCTTCACCGGCACC 369
              | |  || ||||||||  | | ||||  || ||  ||| || ||||     ||
Db        362 GATGGAGGTATCAACCCAGCCTCACTCCGAGGAACGAACACTGGCGTCTGGGTGGGTGTG 421

Qy        370 AACGGACAGGACTACGCACTTCGCGTGCACAACGCGGGCCAGTCAACCGATGGTTTCGCA 429
              |  ||   ||     |      | |   ||  |    | || |       || | |
Db        422 AGTGGTTCAGAGGCATCCGAGGCCCTTAGCAGAGATCCCGAGACGCTTCTGGGCTACAGC 481

Qy        430 CTGACCGGAACCGCCGGCAGCGTCATCTCCGGTCGTATCTCGTACACGTTTGGTTTTGAG 489
              ||   ||   |    |     | ||  ||   ||  |||| | |   || |  ||  |
Db        482 ATGGTGGGCTGCCAGCGTGCAATGATGGCCAACCGGCTCTCTTTCTTCTTCGACTTCAAA 541

Qy        490 GGTCCTGCGGTGTCGGTGGACACGGCTTGTTCCTCGTCGTTGGTGGCTTTGCATCTGGCC 549
              || ||    |  |  ||||||| || || |||||    ||| ||||  | ||    |||
Db        542 GGACCAAGCATTGCCCTGGACACAGCCTGCTCCTCCAGCTTGCTGGCACTACAGAATGCC 601

Qy        550 TGTCAGGCGTTGCGTGCGGGTGAGTGCTCGATGGCGCTTGCCGGGGGTGTGACGGTGATG 609
              |  |||||  | |||   || || ||| |   ||| ||||  || ||  | |   ||
Db        602 TACCAGGCCATCCGTAGTGGGGAATGCCCCGCGGCCCTTGTGGGTGGGATCAACCTGCNN 661

Qy        610 TCGTCTCCG 618
              ||    |||
Db        662 TCTGAGCCG 670
```

RESULT 14
BB865774
LOCUS BB865774 664 bp mRNA linear EST 27-NOV-2001
DEFINITION BB865774 RIKEN full-length enriched, Nullipotent stem cell CRL-2070 NE cDNA Mus musculus cDNA clone G431002L09 5', mRNA sequence.
ACCESSION BB865774
VERSION BB865774.1 GI:17111984
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 664)
AUTHORS Akimura,T., Arakawa,T., Carninci,P., Furuno,M., Hanagaki,T., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T., Imotani,K., Ishii,Y., Ito,M., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Matsuyama,T., Nakamura,M., Nishi,K., Nomura,K., Numasaki,R., Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sakazume,N., Sasaki,D., Sato,K., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki;H., Tagawa,A., Takahashi,F., Takaku-Akahira,S., Tanaka,T., Tomaru,A., Toya,T., Watahiki,A., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y.
TITLE RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura,T., et al. 2001)
JOURNAL Unpublished (2001)
COMMENT Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.go.jp,
URL:http://genome.gsc.riken.go.jp/
Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. . 10 (10), 1617-1630 (2000)
wagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. . 10 (11), 1757-1771 (2000)
Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara,Y. and Hayashizaki,Y.
Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. . 11 (2), 281-289 (2001)
Please visit our web site (http://genome.gsc.riken.go.jp) for further details.
e mouse tissues.

```
FEATURES             Location/Qualifiers
     source          1. .664
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="G431002L09"
                     /clone_lib="RIKEN full-length enriched, Nullipotent stem cell CRL-2070 NE cDNA"
                     /cell_type="Nullipotent stem cell"
                     /cell_line="CRL-2070 NE"
                     /note="pooled cell lines ; (cell_line=CRL-1751 WEHI 164), (cell_line=CRL-2116 JC), (cell_line=RCB-0035 WEHI-3), (cell_line=RCB-0464 Meth-A), (cell_line=RCB-0545 OHTA), (cell_line=RCB-0559 K-1 . F1), (cell_line=RCB-1283 B16 melanoma), (cell_type=B cells, cell_line=CRL-1702 WEHI 231), (cell_type=Leydig cells, cell_line=CRL-2065 MLTC-1), (cell_type=Nullipotent stem cell, cell_line=CRL-2070 NE), (tissue_type=bladder, cell_line=RCB-0544 MBT-2), (tissue_type=bone marrow, cell_type=stroma cell, cell_line= CRL-2028 SR-4987), (tissue_type=colon, cell_line=RCB-0549 Cle-H3), (tissue_type=kidney, cell_line=CCL-142 RAG), (tissue_type=submandibular gland, cell_line=CRL-1734 SCA-9 clone 15), (strain=BALB/C, cell_type=B cells, cell_line=CRL-1669 BCL1 Clone 13.20-3B3), (strain=C3H, tissue_type=brain, cell_line=CRL-1443 BC3H1)"
BASE COUNT      136 a    186 c    206 g    135 t      1 others
ORIGIN
```

Query Match 7.4%; Score 95.2; DB 10; Length 664;
Best Local Similarity 51.8%; Pred. No. 1.5e-08;
Matches 214; Conservative 0; Mismatches 199; Indels 0; Gaps 0;

```
Qy        190 GGATTCCTCTACGACGCAGGCCACTTCGACGCCGAATTCTTCGGCATCAGCCCCCGCGAA 249
              |||   ||  | ||       | | |||||||||   || || ||  ||  ||||    |
Db        245 GGAAAGCTGAAGGATCTCTCCAAGTTCGACGCCTCCTTTTTTGGGGTCCACCCCAAGCAG 304

Qy        250 GCCCTCGCCATGGACCCCCAGCAACGACTCCTCCTCGAAACCGCCTGGGAAACCATCGAA 309
              || | | ||||||||||||||  || || ||  | |||  |  ||  ||| | || |
Db        305 GCACACACAATGGACCCCCAGCTTCGGCTGCTGTTGGAAGTCAGCTATGAAGCAATTGTG 364

Qy        310 CACGCCGGCATCAACCCCCACACCCTCCACGGCACCCCCACCGGAGTCTTCACCGGCACC 369
              | |  || ||||||||  | | ||||  || ||  ||| || ||||     ||
Db        365 GATGGAGGTATCAACCCAGCCTCACTCCGAGGAACGAACACTGGCGTCTGGGTGGGTGTG 424

Qy        370 AACGGACAGGACTACGCACTTCGCGTGCACAACGCGGGCCAGTCAACCGATGGTTTCGCA 429
              |  ||   ||     |      | |   ||  |    | || |       || | |
Db        425 AGTGGTTCAGAGGCATCCGAGGCCCTTAGCAGAGATCCCGAGACGCTTCTGGGCTACAGC 484

Qy        430 CTGACCGGAACCGCCGGCAGCGTCATCTCCGGTCGTATCTCGTACACGTTTGGTTTTGAG 489
              ||   ||   |    |     | ||  ||   ||  |||| | |   || |  ||  |
Db        485 ATGGTGGGCTGCCAGCGTGCAATGATGGCCAACCGGCTCTCTTTCTTCTTCGNACTTCAA 544

Qy        490 GGTCCTGCGGTGTCGGTGGACACGGCTTGTTCCTCGTCGTTGGTGGCTTTGCATCTGGCC 549
              || ||    |  |  ||||||| || || |||||    ||| ||||  | ||    |||
Db        545 GGACCAAGCATTGCCCTGGACACAGCCTGCTCCTCCAGCTTGCTGGCACTACAGAATGCC 604

Qy        550 TGTCAGGCGTTGCGTGCGGGTGAGTGCTCGATGGCGCTTGCCGGGGGGTGTGAC 602
              |  |||||  | |||   || || ||| |   ||| ||||  ||| |  | ||
Db        605 TACCAGGCCATCCGTAGTGGGGAATGCCCCGCGGCCCTTGTGGGGTGGATCAC 657
```

```
RESULT 15
AI532314
LOCUS       AI532314                516 bp    mRNA    linear   EST 19-APR-2001
DEFINITION  SD03777.5prime SD Drosophila melanogaster Schneider L2 cell culture
            pOT2 Drosophila melanogaster cDNA clone SD03777 5prime, mRNA
            sequence.
ACCESSION   AI532314
VERSION     AI532314.1  GI:4446449
KEYWORDS    EST.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 516)
  AUTHORS   Harvey,D., Brokstein,P., Hong,L., Evans-Holm,M., Su,C., Tsang,G.,
            Lewis,S. and Rubin,G.M.
  TITLE     BDGP/HHMI Drosophila EST Project
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Stapleton, M.
            BDGP
            Lawrence Berkeley National Lab
            One Cyclotron Rd, Berkeley, CA 94720, USA
            Fax: 510 486 6798
            Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
            hit genomic sequence AC005554
            Plate: 37  row: G  column: 5
            High quality sequence stop: 514.
FEATURES             Location/Qualifiers
     source          1. .516
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone="SD03777"
                     /clone_lib="SD Drosophila melanogaster Schneider L2 cell
                     culture pOT2"
                     /lab_host="DH5-alpha"
                     /note="Vector: pOT2; Site_1: EcoRI; Site_2: Xho1; Sized
                     fractionated cDNAs were directly ligated into pOT2.
                     Plasmid cDNA library."
BASE COUNT      101 a    163 c    158 g     94 t
ORIGIN

  Query Match             7.3%;  Score 94.6;  DB 9;  Length 516;
  Best Local Similarity   50.9%;  Pred. No. 1.8e-08;
  Matches  257;  Conservative    0;  Mismatches  239;  Indels    9;  Gaps    1;

Qy        458 CCGGTCGTATCTCGTACACGTTTGGTTTTGAGGGTCCTGCGGTGTCGGTGGACACGGCTT 517
              ||  ||| ||||| |||||||| |  ||  |||| ||     |    |||||||| || |
Db          1 CCAATCGCATCTCCTACACGTTCGACTTCAAGGGGCCCAGCTTTATCGTGGACACCGCAT 60

Qy        518 GTTCCTCGTCGTTGGTGGCTTTGCATCTGGCCTGTCAGGCGTTGCGTGCGGGTGAGTGCT 577
              | |||   ||| |||||||  |  | |  ||||    ||   |||| ||||||   ||
Db         61 GCTCCAGCTCGCTGGTGGCCCTTAACCACGCCTTCGCGGACATGCGGGCGGGTCGCTGTG 120

Qy        578 CGATGGCGCTTGCCGGGGGTGTGACGGTGATGTCGTCTCCGGGTGCCTTCGTGGAGTTTT 637
                || || |  | ||| ||||   | ||       ||          | |  ||||
Db        121 ATTACGCCCTGGTTGCGGGCGTGAATCTCATCCTAAAGCCCATCTTTGCCCTCCAGTTCC 180

Qy        638 CGCGGCAGCGGGGTCTGGCCGCGGACGGGCATTGCAAGGCGTTCTCGGCGGCGGCGGACG 697
              |||  |  |     ||  |   |||||    |||||| | |||   |||||||  |||
Db        181 TCAGGCTGGGCATCGTGAGCCACGACGGCTCCTGCAAGACCTTCGATGCGGCGGCCAACG 240

Qy        698 GGACCGGCTGGGGTGAGGGTGTGGGGATGCTGCTGGTGGAGCGGCTCTCCGACGCCCATC 757
              |    | | |||  ||       |   |  ||||| |  || |||     || ||| | |
Db        241 GCTATGCCCGGGCGGACACCTGCGCCGTTGTGCTGCTTCAGAGGCGTAAGGAGGCCAAGC 300

Qy        758 GCAACGGTCACCGTGTCCTGGCCGTGGTGCGTGGCAGTGCGGTCAACCAGGACGGTGCGA 817
              |       | |  |||||   |||  |     |  | |    | ||||
Db        301 GGGTGTACGCCAGCATCCTGAATGTGCGGACCAATACGGACGGATTCAAGGA-------- 352

Qy        818 GCAACGGTCTGACCGCGCCCAACGGGCCGTCCCAGCAGCGTGTCATCCGCCAGGCCCTCG 877
                  | ||  | |||   ||  |||| |      ||||||    |  |     || ||  |
Db        353 -GCAGGGCGTAACCTTCCCAGACGGCCGCATGCAGCAGGCCCTGCTGGAGGAGACCTACA 411

Qy        878 CCAACGCCGGCTTGTCGGCCGGTGATGTCGACGCGGTGGAGGCCCACGGCACCGGCACCA 937
              | |   |||| |      | |  || || | |   |||||||| ||||||| ||||||
Db        412 GCGAGATCGGCCTCAATCCGGACGAGGTGGTCTACGTGGAGGCACACGGCAGCGGCACAC 471

Qy        938 CTTTGGGCGACCCGATCGAGGCCCA 962
              |  |||||||       || ||| |
Db        472 CGGTGGGCGATGACCAGGAAGCCAA 496


Search completed: June 18, 2003, 01:00:23
Job time : 1656.74 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 17, 2003, 11:26:44 ; Search time 1792.37 Seconds
(without alignments)
15003.005 Million cell updates/sec

Title: US-09-914-286-1_COPY_7906_8829
Perfect score: 924
Sequence: 1 accgcattcatctgctccgg.....acccccaaacccacctcctc 924

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		924	100.0	30690	6	E38020	E38020 Avermectin
2		924	100.0	64957	1	AB032367	AB032367 Streptomy
3		919.2	99.5	12381	6	AX006889	AX006889 Sequence
4		917.6	99.3	11096	1	AF275943	AF275943 Streptomy
5		873.4	94.5	31422	6	E38021	E38021 Avermectin
6	c	873.4	94.5	64957	1	AB032367	AB032367 Streptomy
7	c	447.4	48.4	24225	1	SC2C4	AL512902 Streptomy
8		344.6	37.3	32870	1	AF007101	AF007101 Streptomy
9		312.2	33.8	927	1	SCU88833	U88833 Streptomyce
10		308.4	33.4	41097	1	AF016585	AF016585 Streptomy
11		302.2	32.7	20394	1	SNA132222	AJ132222 Streptomy
12		302.2	32.7	20394	6	AX067996	AX067996 Sequence
13	c	302.2	32.7	84985	1	SNA278573	AJ278573 Streptomy
14		300.6	32.5	65140	6	AX211705	AX211705 Sequence
15		300.6	32.5	123580	1	AF263912	AF263912 Streptomy
16		300.6	32.5	125401	6	AX211739	AX211739 Sequence
17		274.2	29.7	113193	1	AF357202	AF357202 Streptomy
18		273.4	29.6	27541	6	AX211706	AX211706 Sequence
19		266.2	28.8	28958	6	AR044578	AR044578 Sequence
20		266.2	28.8	28958	6	I47768	I47768 Sequence 6
21		266.2	28.8	28958	6	I50958	I50958 Sequence 6
22		266.2	28.8	28958	6	I63356	I63356 Sequence 6
23		266.2	28.8	28958	6	I70387	I70387 Sequence 6
24		266.2	28.8	28958	6	I85639	I85639 Sequence 6
25		266.2	28.8	28958	6	I88045	I88045 Sequence 4
26		266.2	28.8	28958	6	I90322	I90322 Sequence 6
27		266.2	28.8	49377	6	I88042	I88042 Sequence 1
28		266.2	28.8	67523	1	SCU24241	U24241 Sorangium c
29		264.4	28.6	56870	2	CEY39B6_3	Continuation (4 of
30		245.6	26.6	87076	9	AC005918	AC005918 Homo sapi
31		244.2	26.4	30000	6	AX250262	AX250262 Sequence
32	c	244	26.4	167254	9	CNS05TDS	AL357093 Human chr
33		243.4	26.3	3030	8	VCA429230	AJ429230 Volvox ca
34	c	240.8	26.1	98734	9	AC013277	AC013277 Homo sapi
35		236.8	25.6	104326	1	AB070940	AB070940 Streptomy
36		235.2	25.5	2203	1	AF202898	AF202898 Streptomy
37	c	234.6	25.4	205691	2	AC087227	AC087227 Mus muscu
38	c	233.8	25.3	200350	2	AC095469	AC095469 Rattus no
39	c	233	25.2	303091	2	AC084799	AC084799 Mus muscu
40		232.8	25.2	78210	1	AB070949	AB070949 Streptomy
41		232.6	25.2	24568	1	AF220951	AF220951 Streptomy
42		232.6	25.2	50937	6	AR159871	AR159871 Sequence
43		231.4	25.0	1080	8	CNS019XP	AL112229 Botrytis
44		229.4	24.8	1080	8	CNS01A6O	AL112552 Botrytis
45	c	228.2	24.7	167390	9	AC007263	AC007263 Homo sapi

ALIGNMENTS

RESULT 1
E38020
LOCUS E38020 30690 bp DNA linear PAT 31-JAN-2002
DEFINITION Avermectin aglycon synthase gene.
ACCESSION E38020
VERSION E38020.1 GI:18626909
KEYWORDS JP 2000245457-A/1.
SOURCE Streptomyces avermitilis.
ORGANISM Streptomyces avermitilis
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE 1 (bases 1 to 30690)
AUTHORS Omura,S. and Ikeda,H.
TITLE Avermectin aglycon synthase gene
JOURNAL Patent: JP 2000245457-A 1 12-SEP-2000;

```
            THE KITASATO INSTITUTE
COMMENT     OS   Streptomyces avermitilis
            PN   JP 2000245457-A/1
            PD   12-SEP-2000
            PF   24-FEB-1999 JP 1999046961
            PR
            PI   SATOSHI OMURA,HARUO IKEDA
            PC   C12N15/00,A61K31/70,C12N1/15,C12N9/88,C12P19/62, PC
            C12Q1/68//C07D493/22,
            PC   (C12N1/15,C12R1:465),(C12N9/88,C12R1:465),C12N15/00 CC
            FH   Key             Location/Qualifiers
            FT   CDS             (1). .(11916)
            FT   CDS             (11971). .(30687).
FEATURES             Location/Qualifiers
     source          1. .30690
                     /organism="Streptomyces avermitilis"
                     /db_xref="taxon:33903"
BASE COUNT     5355 a  12455 c   8617 g   4263 t
ORIGIN

  Query Match             100.0%;  Score 924;  DB 6;  Length 30690;
  Best Local Similarity   100.0%;  Pred. No. 4.7e-89;
  Matches  924;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ACCGCATTCATCTGCTCCGGACAGGGCACCCAACGCCCCGGCATGGCCCACGGCCTCTAC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7906 ACCGCATTCATCTGCTCCGGACAGGGCACCCAACGCCCCGGCATGGCCCACGGCCTCTAC 7965

Qy         61 CACACCCACCCCGTCTTCGCCGCCGCACTCAACGACATCTGCACCCACCTCGACCCCCAC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7966 CACACCCACCCCGTCTTCGCCGCCGCACTCAACGACATCTGCACCCACCTCGACCCCCAC 8025

Qy        121 CTCGACCACCCCCTCCTCCCCCTCCTCACCCAAAACGACAACGACAACGAGGACGCGGCC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8026 CTCGACCACCCCCTCCTCCCCCTCCTCACCCAAAACGACAACGACAACGAGGACGCGGCC 8085

Qy        181 GCACTGCTCCAGCAGACCCGCTACGCCCAGCCCGCCCTCTTCGCCTTCCAGGTCGCCCTC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8086 GCACTGCTCCAGCAGACCCGCTACGCCCAGCCCGCCCTCTTCGCCTTCCAGGTCGCCCTC 8145

Qy        241 CACCGCCTCCTCACCGACGGCTACCACATCACCCCCCACTACTACGCCGGACACTCCCTC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8146 CACCGCCTCCTCACCGACGGCTACCACATCACCCCCCACTACTACGCCGGACACTCCCTC 8205

Qy        301 GGCGAAATCACCGCCGCCCCACCTCGCCGGCATCCTCACCCTCACCGACGCCACCACCCTC 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8206 GGCGAAATCACCGCCGCCCCACCTCGCCGGCATCCTCACCCTCACCGACGCCACCACCCTC 8265

Qy        361 ATCACCCAACGCGCCACCCTCATGCAAACCATGCCCCCCGGCACCATGACCACCCTCCAC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8266 ATCACCCAACGCGCCACCCTCATGCAAACCATGCCCCCCGGCACCATGACCACCCTCCAC 8325

Qy        421 ACCACCCCCCACCACATCACCCACCACCTCACCGCCCACGAAAACGACCTCGCCATCGCC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8326 ACCACCCCCCACCACATCACCCACCACCTCACCGCCCACGAAAACGACCTCGCCATCGCC 8385

Qy        481 GCCATCAACACCCCCACCTCCCTCGTCATCAGCGGCACCCCCCACACCGTCCAACACATC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8386 GCCATCAACACCCCCACCTCCCTCGTCATCAGCGGCACCCCCCACACCGTCCAACACATC 8445

Qy        541 ACCACCCTCTGCCAACAACAAGGCATCAAAACCAAAACCCTCCCCACCAACCACGCCTTC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8446 ACCACCCTCTGCCAACAACAAGGCATCAAAACCAAAACCCTCCCCACCAACCACGCCTTC 8505

Qy        601 CACTCCCCCCACACCAACCCCATCCTCAACCAACTCCACCAGCACACCCAAACCCTCACC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8506 CACTCCCCCCACACCAACCCCATCCTCAACCAACTCCACCAGCACACCCAAACCCTCACC 8565

Qy        661 TACCACCCACCCCACACCCCCCTCATCACCGCCAACACCCCACCCGACCAACTCCTCACC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8566 TACCACCCACCCCACACCCCCCTCATCACCGCCAACACCCCACCCGACCAACTCCTCACC 8625

Qy        721 CCCCACTACTGGACCCAACAAGCCCGCAACACCGTCGACTACGCCACCACCACCCAAACC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8626 CCCCACTACTGGACCCAACAAGCCCGCAACACCGTCGACTACGCCACCACCACCCAAACC 8685

Qy        781 CTCCACCAACACGGCGTCACCACCTACATCGAACTCGGACCCGACAACACCCTCACCACC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8686 CTCCACCAACACGGCGTCACCACCTACATCGAACTCGGACCCGACAACACCCTCACCACC 8745

Qy        841 CTCACCCACCACAACCTCCCCAACCCCCCCACCACCACCCTCACCCTCACCCACCCCCAC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8746 CTCACCCACCACAACCTCCCCAACCCCCCCACCACCACCCTCACCCTCACCCACCCCCAC 8805

Qy        901 CACCACCCCCAAACCCACCTCCTC 924
              |||||||||||||||||||||||||
Db       8806 CACCACCCCCAAACCCACCTCCTC 8829


RESULT 2
AB032367
LOCUS       AB032367               64957 bp    DNA     linear   BCT 18-SEP-1999
DEFINITION  Streptomyces avermitilis polyketide synthase gene cluster (aveA1,
            aveA2, aveA3, aveA4) and aveC, aveE genes, complete cds.
ACCESSION   AB032367
VERSION     AB032367.1  GI:5902890
KEYWORDS    AveC; cytochrome P450 hydroxylase; type I polyketide synthase AVES
            4; type I polyketide synthase AVES 3; type I polyketide synthase
            AVES 2; type I polyketide synthase AVES 1.
SOURCE      Streptomyces avermitilis DNA.
  ORGANISM  Streptomyces avermitilis
            Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
            Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1  (bases 1 to 64957)
  AUTHORS   Ikeda,H., Nonomiya,T., Usami,M., Ohta,T. and Omura,S.
  TITLE     Organization of the biosynthetic gene cluster for the polyketide
            anthelmintic macrolide avermectin in Streptomyces avermitilis
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9509-9514 (1999)
  MEDLINE   99380548
REFERENCE   2  (bases 1 to 64957)
  AUTHORS   Ikeda,H., Nonomiya,T., Usami,M., Ohta,T. and Omura,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-SEP-1999) Haruo Ikeda, School of Pharmaceutical
            Sciences, Kitasato University, Microbial chemistry; 5-9-1
            Shirokane, Minato-ku, Tokyo 108-8641, Japan
            (E-mail:ikeda@mc.pharm.kitasato-u.ac.jp, Tel:+81-3-5791-6242,
            Fax:+81-3-3444-6197)
FEATURES             Location/Qualifiers
     source          1. .64957
                     /organism="Streptomyces avermitilis"
                     /db_xref="taxon:33903"
     gene            101. .12019
                     /gene="aveA1"
     CDS             101. .12019
                     /gene="aveA1"
                     /note="contains a loading module, Ave module 1 and module
                     2
                     multifunctional polyketide synthase"
                     /codon_start=1
                     /transl_table=11
                     /product="type I polyketide synthase AVES 1"
                     /protein_id="BAA84474.1"
                     /db_xref="GI:5902891"
                     /translation="MQRMDGGEEPRPAAGEVLGVADEADGGVVFVFPGQGPQWPGMGR
                     ELLDASDVFRESVRACEAAFAPYVDWSVEQVLRDSPDAPGLDRVDVVQPTLFAVMISL
                     AALWRSQGVEPCAVLGHSLGEIAAAHVSGGLSLADAARVVTLWSQAQTTLAGTGALVS
                     VAATPDELLPRIAPWTEDNPARLAVAAVNGPRSTVVSGAREAVADLVADLTAAQVRTR
                     MIPVDVPAHSPLMYAIEERVVSGLLPITPRPSRIPFHSSVTGGRLDTRELDAAYWYRN
                     MSSTVRFEPAARLLLQQGPKTFVEMSPHPVLTMGLQELAPDLGDTTGTADTVIMGTLR
                     RGQGTLDHFLTSLAQLRGHGETSATTVLSARLTALSPTQQQSLLLDLVRAHTMAVLND
                     DGNERTASDAGPSASFAHLGFDSVMGVELRNRLSKATGLRLPVTLIFDHTTPAAVAAR
                     LRTAALGHLDEDTAPVPDSPSGHGGTAAADDPIAIIGMACRFPGGVRSPKDLWELAAS
                     GGDAIGPFPTDRGWPTEQRHAQDPTQPGTFYPQGGGFLHDAAHFDAGFFGISPREALA
                     MDPQQRLLLETSWEAFERAGIDPLSVRGSRTGVFAGALSFDYGPRMDTASSEGAADVE
                     GHILTGTTGSVLSGRIAYSFGLEGPAITVDTGCSASLVTLHLACQSLRSGECTLALAG
                     GVSVMSTLGMFIEFSRQRGLSVDGRCKAYSAAADGTGWGEGVGMLLVERLSDAVRLGH
```

RVLAVVRGSAVNQDGASNGLTAPNGPAQERVIRQALANAGLSVADVDVVEGHGTGTTL
GDPIEAQALLATYGQRAGDRPLWLGSLKSNIGHTMAAAGVGGVIKMVMALREGVLPRT
LHVDKPSPQVDWSAGAVRLLTEAVPWPGDAAGRLRRAGVSSFGIGGTNAHVILEEAPA
AGGCVAGGGVLEGAPGLAISVAESVAAPVAVSAPVAESVPVPVPVPVPVSARSEAG
LRAQAEALRQYVAVRPDVSLADVGAGLACGRAVLEHRAVVLAADREELVQGLGALAAG
EPDRRVTTGHAPGGDRGGVVFVFPGQGGQWAGMGVRLLASSPVFARRMQACEEALAPW
VDWSVVDILRRDAGDAVWERADVVQPVLFSVMVSLAALWRSYGIEPDAVLGHSQGEIA
AAHVCGALSLKDAAKTVALRSRALAAVRGRGGMASVPLPAQEVEQLIGERWAGRLWVA
AVNGPRSTAVSGDAEAVDEVLAYCAGTGVRARRIPVDYASHCPHVQPLREELLELLGD
ISPQPSGVPFFSTVEGTWLDTTTLDAAYWYRNLHQPVRFSDAVQALADDGHRVFVEVS
PHPTLVPAIEDTTEDTAEDVTAIGSLRRGDNDTRRFLTALAHTHTTGIGTPTTWHHHY
THHHTHPHPHTHLDLPTYPFQHQHYWLESSQPGAGSGSGAGAGSGAGSGRAGTAGGTA
EVESRFWDAVARQDLETVATTLAVPPSAGLDTVVPALSAWHRHQHDQARINTWTYQET
WKPLTLPTTHQPHQTWLIAIPETQTHHPHITNILTNLHHHGITPIPLTLNHTHTNPQH
LHHTLHHTRQQAQNHTTGAITGLLSLLALDETPHPHHPHTPTGTLLNLTLTQTHTQTH
PPTPLWYATTNATTTHPNDPLTHPTQAQTWGLARTTLLEHPTHTAGIIDLPTTPTPHT
LQHLTQTLTQPHHQTQLAIRTTGTHTRRLTPTTLTPTHQPPTPTPHGTTLITGGTGAL
ATHLTHHLTTHQPTQHLLLTSRTGPHTPHAQHLTTQLQQKGIHLTITTCDTSNPDQLQ
QLLNTIPPQHPLTTVIHTAGILDDATLTNLTPTQLNNVLRAKAHSAHLLHQLTQHTPL
TAFVLYSSAAATFGAPGQANYAAANAYLDALAHHRHTHHLPATSIAWGTWQGNGLADS
DKARAYLDRRGFRPMSPELATAAVTQAIADTERPYVVIADIDWSKIEHTSQTSDLVSA
ARAREPAVQRPTPPAELHKTLAHQTSADQRAALLELVRDHVAAVLRHADPKAIAPDQS
FRALGFDSLTAVEFRNLLIKATGLRLPVSLVFDHPTPAKLAVHLQNQLRGTAAESAPS
AAAVTAEASVTEPIAIVGMACRFPGGVTSADDFWDLISSEQDAIGGFPTDRGWDLDTL
YDPDPDHPGTCYTRNGGFLYDAGHFDAEFFGISPREALAMDPQQRLLLETAWETIEHA
GINPHTLHGTPTGVFTGTNGQDYALRVHNAGQSTDGFALTGTAGSVISGRISYTFGFE
GPAVSVDTACSSSLVALHLACQALRAGECSMALAGGVTVMSSPGAFVEFSRQRGLAAD
GHCKAFSAAADGTGWGEGVGMLLVERLSDAHRNGHRVLAVVRGSAVNQDGASNGLTAP
NGPSQQRVIRQALANAGLSAGDVDAVEAHGTGTTLGDPIEAQALLATYGQDRAGEGPL
WLGSVKSNVGHTQAAAGVAGVIKMVMALRHGLLPRTLHVDEPSPHVDWSAGAVQLLTE
TVPWPGGEGRLRRAGVSSFGVSGTNAHVILEEAPADDVPGGPPAGEGDAGSDDEAAAG
SPGVWPWLVSAKSQPALRAQAQALHAHLTDHPGLDLADVGYTLAHARAVFDHRATLIA
ADRDTFLQALQALAAGEPHPAVIHSSAPGGTGTGEAAGKTAFICSGQGTQRPGMAHGL
YHTHPVFAAALNDICTHLDPHLDHPLLPLLTQNDNDNEDAAALLQQTRYAQPALFAFQ
VALHRLLTDGYHITPHYYAGHSLGEITAAHLAGILTLTDATTLITQRATLMQTMPPGT
MTTLHTTPHHITHHLTAHENDLAIAAINTPTSLVISGTPHTVQHITTLCQQQGIKTKT
LPTNHAFHSPHTNPILNQLHQHTQTLTYHPPHTPLITANTPPDQLLTPHYWTQQARNT
VDYATTTQTLHQHGVTTYIELGPDNTLTTLTHHNLPNPPTTTLTLTHPHHHPQTHLLT
NLAKTTTTWHPHHYTHHDNQPHTHTHLDLPTYPFQHHHYWLESTQPGAGNVSAAGLDP
TEHPLLGATLELATDGGALLAGRLSLRSHPWLADHAVGGTVLLSGATFLELALHAGTY
VGCDRVDELTLHAPLVVPVDGGVSVQVGVAAADGEGRRLVSVYARGGSACGGGASGG
VWTCHASGVLVEAAAGGVVVDGLAGVWPPRGAVAVDVDGVRDRLAGAGCVLGPVFSGL
RAVWRDGGDLLAEVCLPEEAWGDAAGFGLHPALLDGVVQPLSVLLPGGTGFGEGAGFG
EGVRVPAVWGGVSLHRAGVTGVRVRVSAVGRGGGREAVSVVVGDEAGVPVASVDRLEL
RPVDMGQLRAVSVSAGRRGSLYAVQWAEVGPVPVCGQAWAWHEDVGESGGGPVPGVVV
LRCPDAGAGGGGGGGGGGVGEVVGGVLGVVQGWLGLERFAGSRLVVVTRGAVVAGPE
DGPVDVVGASVWGLVRSAQAEHPDRFVLLDLDTDTGTDLDTGAGAGWGVDGGRVAAVV
ACGEPQLAVRGERLLAARLKRLESSGDVPAQRSGDTRARRSDVPAQRSGGVPARRSVD
VSGREVLPWLSGGSVLVTGGTGVLGAAVARHLAGVCGVRDLLLVSRRGPDAPGAEGLR
AELAALGAEVRIVACDVGERREVVRLLEGVPAGCPLTGVVHAAGVLDDATIASLTPER
LGTVFAAKVDAALLLDELTRGMELSAFVLFSSAAGILGSAGQGNYAAANAALDALAYR
RRAAGLPGVSLAWGLWEEASGMTGHLAGTDHRRIIRSGLHPMSTPDALALFDAALALD
RPVLLPADLRPAPPLPPLLQDLLPATRRRTTRTTTTGGADNGAQLHARLAGQTHEQQH
TTLLALVRSHIATVLGHTTPDTIPPDRAFRDLGFDSLTAVELRNRLSRTTGLRLPTTL
AFDHPNPTTLTHHLHTQLQPQPDNAVAPVLAELDKLESALSALDKTDSASERVTLRLK
SLMLRWNAPQHPTAESADDDEKFTSATEAEIFKFIDNDLGLS"

gene 12071. .30790
/gene="aveA2"

CDS 12071. .30790
/gene="aveA2"
/note="contains Ave module 3, module 4, module 5 and module 6
multifunctional polyketide synthase"
/codon_start=1
/transl_table=11
/product="type I polyketide synthase AVES 2"
/protein_id="BAA84475.1"
/db_xref="GI:5902892"
/translation="MQLANEAKLLEYLKRVTADLDRTRRRLYEVVEREQEPIAIVGMA
CRYPGGATSPTRLWHLVKSQTDAIGEFPTDRGWNLEQLYDPDPDRSGTSYTRSGGFLY
DAGDFDAAFFELSPREALAMDPQQRLLLETTWETFEQGGIDPRSMRGSRTGVFVGINP
EDYTTGYTHQPSNAVEGYLLTGSAASIASGRISYNFGLEGPAITIDTACSSSLVALHL
ACQALRSGECTMALAGGASVMATPFVFTEFSRQRGLAADGRCKAFSAAADGTGWSEGV
GMLLVERLSDARRNGHRVLAVVRGSAVNQDGASNGLTAPNGRSQVKVIRQALANAHLS
PADVDAVEAHGTGTTLGDPIEAQALVEAYGQDRPNGRPLWLGTLKSNIGHSMAAAGVG
GVIKMVMALRNGLLPRTLHVDEPSPHVDWSAGAVQLLTETVPWPGGEGRLRRAGVSSF
GVSGTNAHVILEEAPAHNIPSDTPADDVPGESAADEDAGSGDEAAAGSPGVWPWLVSA
KSQPALRAQAQALHAHLTDHPGLDLADVGYTLAHARAVFDHRATLIAADRDTFLQALQ
ALAAGEPHPAVIHSSAPGGTGTGEAAGKTAFICSGQGTQRPGMAHGLYHTHPVFAAAL
NDICTHLDPHLDHPLLPLLTQDPNTQDTTTLEEAAALLQQTRYAQPALFAFQVALHRL
LTDGYHITPHYYAGHSLGEITAAHLAGILTLTDATTLITQRATLMQTMPPGTMTTLHT
TPHHITHHLTAHENDLAIAAINTPTSLVISGTPHTVQHITTLCQQQGIKTKTLPTNHA
FHSPHTNPILNQLHQHTQTLTYHPPHTPLITANTPPDQLLTPHYWTQQARNTVDYATT
TQTLHQHGVTTYIELGPDNTLTTLTHHNLPNTPTTTLTLTHPHHHPQTHLLTNLAKTT
TTWHPHHYTHHHNQPHTHTHLDLPTYPFQHHHYWLQPPGKPSDPSPSEGREQATTPST
PLRDVLVGKSPQERDEELLRLVRTHAAAVLGHATPEVIVPNKAFKELGFDSLAAIQLR
NRLLADVDLPLPATLIFDYPTPMALCQFLRAAIVGADTGTTTRLPLTAVPADEPIAIV
GMACRYPGDVRTVDDLWQVVSGGHDAIGGFPTNRGWDLDTLYNPDPDHHGTSYTRSGG
FLYDAGNFDPDFFGISPREALAMDPQQRLLLETAWESIEHACINPDSLRGTPTGVFAG
LTYHDYAARFPTAPAGFEGYLGHGSAGSIASGRVAYALGLEGPALTVDTACSSSLVAL
HLACQALRSGECSMALAGGVTVMSTPAGFVEFSRQRGLAVDGRCKAFSAAADGTGWGE
GVGMLLVERLSDARRLGHRILAVVRGSAVNQDGASNGLTAPNGPSQERVIRLALANAD
LTPADVDAVEAHGTGTTLGDPIEAQALLATYGQDRPGNEPLWLGSMKSNIGHAQAAAG
VGGVIKMVMALRNGLLPRTLHVDEPSPHVDWSAGAVQLLTETVPWPGGEGRLRRAGVS
SFGVSGTNAHVILEEAPAHNIPSDTPADDAPGEAAADDVPGEAAGDDAGTGGEATGPA
AGSPGVWPWLVSAKSQPALRAQAQALHAHLTDHPGLDLADVGYTLAHARAVFDHRATL
IAADRDTFLQALQALAAGEPHPAVIHSSAPGGTGTGEAAGKTAPICSGQGTQRPGMAH
GLYHTHPVFAAALNDICTHLDPHLDHPLLPLLTQDPNTQDTTTLEEAAALLQQTPYAQ
PALFAFQVALHRLLTDGYHITPHYYAGHSLGEITAAHLAGILTLTDATTLITQRATLM
QTMPPGTMTTLHTTPHHITHHLTAHENDLAIAAINTPTSLVISGTPHTVQHITTLCQQ
QGIKTKTLPTNHAFHSPHTNPILNQLHQHTQTLTYHPPHTPLITANTPPDQLLTPHYW
TQQARNTVDYATTTQTLHQHGVTTYIELGPDNTLTTLTHHNLPNTPTTTLTLTHPHHH
PQTHLLTNLAKTTTTWHPHHYTHHHNQPHTHTHLDLPTYPFQHQHYWLESTQPGAGSG
SGSGSGRAGTAGGTAEVESRFWDAVARQDLETVATTLAVPPSAGLDTVVPALSAWHRH
QHDQARINTWTYQETWKPLTLPTTHQPHQTWLIAIPETQTHHPHITNILTNLHHHGIT
PIPLTLNHTHTNPQHLHHTRQQAQNHTTGPITGLLSLLALDETPHPHHPHTPTGTLLN
LTLTQTHTQTHPPTPLWYATTNATTTHPNDPLTHPTQAQTWGLARTTLLEHPTHTAGI
IDLPTTPTPHTLHHLTQTLTQPHHQTQLAIRTTGTHTRRLTPTTLTPTHQPPTPTPHG
TTLITGGTGALATHLTHHLTTHQPTQHLLLTSRTGPHTPHAQHLTTQLQQKGIHLTIT
TCDTSNPDQLQQLLNTIPPQHPLTTVIHTAGILDDATLTNLTPTQLNNVLRAKAHSAH
LLHQLTQHTPLNAFVLYSSAAATFGAPGQANYAAANAYLDALAHHRHTHHLPATSIAW
GTWQGNGLATGQVSEHLRRRGMFAMPPELAVTAVDGAIASGRPSLLVADIDWKKLGPV
LSSKSSVLLEDLPQAQGTEEARSTVEQTESTNLRQLLMGRSRSEQEEELLSLVRIHSA
AVLGRDDSEAIPPGRLFRDLGFDSLAAVELRNHLAAQTELALPTTLVFDYPSPTKLAQ
FLLSEIAEFQPDNSTPLPRPRAELDEPIAIVGMACRFPGGVTSADDFWDLISSEQDAI
GGFPTDRGWDLDTLYDPDPDHPGTCYTRNGGFLYDAGHFDAEFFGISPREALAMDPQQ
RLLLETAWETIEHAGINPHTLHGTPTGVFTGTNGQDHAAHIRQAPSGTEGFVLTGAAT
SIASGRISYILGLEGPAVTLDTACSSSLVALHLACQSLRSGECTMALAGGATVMTTPI
TFTEFARQRGLAPDGRCKAFSAAADGTGWGEGVGMLLVERLSDARRNGHRVLAVVRGS
AVNQDGASNGLTAPNGPSQQRVIRQALANADLTPADVDAVEAHGTGTTLGDPIEAQAI
LATYGQDRPGNGPLWLGSVKSNVGHTQAAAGVAGVIKMVMALRHRTLPPTLHADEPSP
HVDWSAGAVQLLTETVPWPGGEGRPRRAGVSSFGVSGTNAHVILEEAPADDVPGGPPA
DEDAGSGEEAAAGSPGVWPWLVSAKSQPALRAQAQALHAHLTDHPGLDLADVGYTLAH
ARAVFDHRATLIAADRDTFLQALQALAAGEPHPAVIHSSAPGGTGTGEAAGKTAFICS
GQGTQRPGMAHGLYHTHPVFAAALNDICTHLDPHLDHPLLPLLTQNDNDNDNEDAAAL
LQQTPYAQPALFAFQVALHRLLTDGYHITPHYYAGHSLGEITAAHLAGILTLTDATTL
ITQRATLMQTMPPGTMTTLHTTPHHITHHLTAHENDLAIAAINTPTSLVISGTPHTVQ
HITTLCQQQGIKTKTLPTNHAFHSPHTNPILNQLHQHTQTLTYHPPHTPLITANTPPD
QLLTPHYWTQQARNTVDYATTTQTLHQHGVTTYIELGPDNTLTTLTHHNLPNTPTTTL
TLTHPHHHPQTHLLTNLAKTTTTWHPHHYTHHHNQPHTHTHLDLPTYPFQHHHYWLEL
PSAQTSPGQRRSRRSAPDTAESEFWDAVNEEDLQSLAETLDIDASALDTVVPALSAWH
RHQHDQARINTWTYQETWKPLTLPTTHQPHQTWLIAIPETQTHHPHITNILTNLHHHG
ITPIPLTVNHTHTNPQHLHHTLHHTRQQAQNHTTGPITGLLSLLALDETPHPHHPHTP
TGTLLNLTLPQTHTQTHPPTPLWYATTNATTTHPNDPLTHPTQAQTWGLARTTLLEHP
THTAGIIDLPTTPTPHTLHHLTQTLTQPHHQTQLAIRTTGTHTRRLTPTTLTPTHQPP
TPTPHGTTLITGGTGALATHLTHHLTTHQPTQHLLLTSRTGPHTPHAQHLTTQLQQKG
IHLTITTCDTSNPDQLQQLLNTIPPQHPLTTVIHTAGVNLFAPVSETDAESFSSVTAA
KATGAAILHELLLDHETLEHFILFSSGAGAWGSGNQCAYSAANAYLDALATHRQTHGL
PGASIAWGPWAGKGMSAGDAAHGYLEKRGILPMEPRMALAAFHRARAQRPNSNLIIAD
IDWERFVPAFTARRHSPLIEDIPEVRQAAQELEAAASTAKTTTAQPIATSLRERLARL
TSSKQNQVLLGLIRTGICTVLGLRNPEGIEDQRAFRDLGFDSLTSAQFSKELAKETGL
PLPPSLVFDYPTPQECAAHLRTQLVDLDDEEDAALSNALPQVAHRRTVEDEPIAIIGM
ACRFPGGVRSADDLWELLASGKDAIGVFPTDRGWDLDTLYDPDPDHPGTCYTRNGGFL
YGAGHFDAEFFGISPREALAMDPQQRLLLETAWETIEHAGINPHTLHGTPTGVFAGIN
AQDHAAHIRQSRDVETIEGYALTGSSGSVASGRVAYTLGLEGPAVSVDTACSSSLVAL

```
  Query Match            100.0%;  Score 924;  DB 1;  Length 64957;
  Best Local Similarity  100.0%;  Pred. No. 4e-89;
  Matches  924;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ACCGCATTCATCTGCTCCGGACAGGGCACCCAACGCCCCGGCATGGCCCACGGCCTCTAC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8006 ACCGCATTCATCTGCTCCGGACAGGGCACCCAACGCCCCGGCATGGCCCACGGCCTCTAC 8065

Qy         61 CACACCCACCCCGTCTTCGCCGCCGCACTCAACGACATCTGCACCCACCTCGACCCCCAC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8066 CACACCCACCCCGTCTTCGCCGCCGCACTCAACGACATCTGCACCCACCTCGACCCCCAC 8125

Qy        121 CTCGACCACCCCCTCCTCCCCCTCCTCACCCAAAACGACAACGACAACGAGGACGCGGCC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8126 CTCGACCACCCCCTCCTCCCCCTCCTCACCCAAAACGACAACGACAACGAGGACGCGGCC 8185

Qy        181 GCACTGCTCCAGCAGACCCGCTACGCCCAGCCCGCCCTCTTCGCCTTCCAGGTCGCCCTC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8186 GCACTGCTCCAGCAGACCCGCTACGCCCAGCCCGCCCTCTTCGCCTTCCAGGTCGCCCTC 8245

Qy        241 CACCGCCTCCTCACCGACGGCTACCACATCACCCCCCACTACTACGCCGGACACTCCCTC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8246 CACCGCCTCCTCACCGACGGCTACCACATCACCCCCCACTACTACGCCGGACACTCCCTC 8305

Qy        301 GGCGAAATCACCGCCGCCCACCTCGCCGGCATCCTCACCCTCACCGACGCCACCACCCTC 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8306 GGCGAAATCACCGCCGCCCACCTCGCCGGCATCCTCACCCTCACCGACGCCACCACCCTC 8365

Qy        361 ATCACCCAACGCGCCACCCTCATGCAAACCATGCCCCCCGGCACCATGACCACCCTCCAC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8366 ATCACCCAACGCGCCACCCTCATGCAAACCATGCCCCCCGGCACCATGACCACCCTCCAC 8425

Qy        421 ACCACCCCCCACCACATCACCCACCACCTCACCGCCCACGAAAACGACCTCGCCATCGCC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8426 ACCACCCCCCACCACATCACCCACCACCTCACCGCCCACGAAAACGACCTCGCCATCGCC 8485

Qy        481 GCCATCAACACCCCCACCTCCCTCGTCATCAGCGGCACCCCCCACACCGTCCAACACATC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8486 GCCATCAACACCCCCACCTCCCTCGTCATCAGCGGCACCCCCCACACCGTCCAACACATC 8545

Qy        541 ACCACCCTCTGCCAACAACAAGGCATCAAAACCAAAACCCTCCCCACCAACCACGCCTTC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8546 ACCACCCTCTGCCAACAACAAGGCATCAAAACCAAAACCCTCCCCACCAACCACGCCTTC 8605

Qy        601 CACTCCCCCCACACCAACCCCATCCTCAACCAACTCCACCAGCACACCCAAACCCTCACC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8606 CACTCCCCCCACACCAACCCCATCCTCAACCAACTCCACCAGCACACCCAAACCCTCACC 8665

Qy        661 TACCACCCACCCCACACCCCCCTCATCACCGCCAACACCCCACCCGACCAACTCCTCACC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8666 TACCACCCACCCCACACCCCCCTCATCACCGCCAACACCCCACCCGACCAACTCCTCACC 8725

Qy        721 CCCCACTACTGGACCCAACAAGCCCGCAACACCGTCGACTACGCCACCACCACCCAAACC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8726 CCCCACTACTGGACCCAACAAGCCCGCAACACCGTCGACTACGCCACCACCACCCAAACC 8785

Qy        781 CTCCACCAACACGGCGTCACCACCTACATCGAACTCGGACCCGACAACACCCTCACCACC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8786 CTCCACCAACACGGCGTCACCACCTACATCGAACTCGGACCCGACAACACCCTCACCACC 8845

Qy        841 CTCACCCACCACAACCTCCCCAACCCCCCCACCACCACCCTCACCCTCACCCACCCCCAC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8846 CTCACCCACCACAACCTCCCCAACCCCCCCACCACCACCCTCACCCTCACCCACCCCCAC 8905

Qy        901 CACCACCCCCAAACCCACCTCCTC 924
              |||||||||||||||||||||||||
Db       8906 CACCACCCCCAAACCCACCTCCTC 8929


RESULT 3
AX006889
LOCUS       AX006889               12381 bp    DNA     linear   PAT 06-SEP-2000
DEFINITION  Sequence 1 from Patent WO0001827.
ACCESSION   AX006889
VERSION     AX006889.1  GI:9994904
KEYWORDS    .
SOURCE      Streptomyces avermitilis.
  ORGANISM  Streptomyces avermitilis
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1  (bases 1 to 12381)
  AUTHORS   Kellenberger,J.L., Leadlay,P.F., Staunton,J., Mcarthur,H.A. and
            Stutzman-Engwall,K.J.
  TITLE     Polyketides, their preparation, and materials for use therein
  JOURNAL   Patent: WO 0001827-A 1 13-JAN-2000;
            KELLENBERGER JOHANNES LAURENZ (CH); LEADLAY PETER FRANCIS (GB);
            STAUNTON JAMES (GB); BIOTICA TECH LTD (GB); PFIZER (US); MCARTHUR
            HAMISH ALASTAIR IRVIN (US); STUTZMAN ENGWALL KIM JONELLE (US)
FEATURES             Location/Qualifiers
     source          1. .12381
                     /organism="Streptomyces avermitilis"
                     /db_xref="taxon:33903"
BASE COUNT     1884 a   4561 c   4005 g   1931 t
ORIGIN

  Query Match            99.5%;  Score 919.2;  DB 6;  Length 12381;
  Best Local Similarity  99.7%;  Pred. No. 1.8e-88;
  Matches  921;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          1 ACCGCATTCATCTGCTCCGGACAGGGCACCCAACGCCCCGGCATGGCCCACGGCCTCTAC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8723 ACCGCATTCATCTGCTCCGGACAGGGCACCCAACGCCCCGGCATGGCCCACGGCCTCTAC 8782

Qy         61 CACACCCACCCCGTCTTCGCCGCCGCACTCAACGACATCTGCACCCACCTCGACCCCCAC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8783 CACACCCACCCCGTCTTCGCCGCCGCACTCAACGACATCTGCACCCACCTCGACCCCCAC 8842

Qy        121 CTCGACCACCCCCTCCTCCCCCTCCTCACCCAAAACGACAACGACAACGAGGACGCGGCC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8843 CTCGACCACCCCCTCCTCCCCCTCCTCACCCAAAACGACAACGACAACGAGGACGCGGCC 8902

Qy        181 GCACTGCTCCAGCAGACCCGCTACGCCCAGCCCGCCCTCTTCGCCTTCCAGGTCGCCCTC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8903 GCACTGCTCCAGCAGACCCGCTACGCCCAGCCCGCCCTCTTCGCCTTCCAGGTCGCCCTC 8962

Qy        241 CACCGCCTCCTCACCGACGGCTACCACATCACCCCCCACTACTACGCCGGACACTCCCTC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8963 CACCGCCTCCTCACCGACGGCTACCACATCACCCCCCACTACTACGCCGGACACTCCCTC 9022

Qy        301 GGCGAAATCACCGCCGCCCACCTCGCCGGCATCCTCACCCTCACCGACGCCACCACCCTC 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9023 GGCGAAATCACCGCCGCCCACCTCGCCGGCATCCTCACCCTCACCGACGCCACCACCCTC 9082

Qy        361 ATCACCCAACGCGCCACCCTCATGCAAACCATGCCCCCCGGCACCATGACCACCCTCCAC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9083 ATCACCCAACGCGCCACCCTCATGCAAACCATGCCCCCCGGCACCATGACCACCCTCCAC 9142

Qy        421 ACCACCCCCCACCACATCACCCACCACCTCACCGCCCACGAAAACGACCTCGCCATCGCC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9143 ACCACCCCCCACCACATCACCCACCACCTCACCGCCCACGAAAACGACCTCGCCATCGCC 9202

Qy        481 GCCATCAACACCCCCACCTCCCTCGTCATCAGCGGCACCCCCCACACCGTCCAACACATC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9203 GCCATCAACACCCCCACCTCCCTCGTCATCAGCGGCACCCCCCACACCGTCCAACACATC 9262

Qy        541 ACCACCCTCTGCCAACAACAAGGCATCAAAACCAAAACCCTCCCCACCAACCACGCCTTC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9263 ACCACCCTCTGCCAACAACAAGGCATCAAAACCAAAACCCTCCCCACCAACCACGCCTTC 9322

Qy        601 CACTCCCCCCACACCAACCCCATCCTCAACCAACTCCACCAGCACACCCAAACCCTCACC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9323 CACTCCCCCCACACCAACCCCATCCTCAACCAACTCCACCAGCACACCCAAACCCTCACC 9382

Qy        661 TACCACCCACCCCACACCCCCCTCATCACCGCCAACACCCCACCCGACCAACTCCTCACC 720
              |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db       9383 TACCACCCACCCCACACCCCCCTCATCACCGACAACACCCCACCCGACCAACTCCTCACC 9442
```

```
Qy        721 CCCCACTACTGGACCCAACAAGCCCGCAACACCGTCGACTACGCCACCACCACCCAAACC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9443 CCCCACTACTGGACCCAACAAGCCCGCAACACCGTCGACTACGCCACCACCACCCAAACC 9502

Qy        781 CTCCACCAACACGGCGTCACCACCTACATCGAACTCGGACCCGACAACACCCTCACCACC 840
              ||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db       9503 CTCCACCAACACGGCGTCACCACCTACATCGATCTCGGACCCGACAACACCCTCACCACC 9562

Qy        841 CTCACCCACCACAACCTCCCCAACCCCCCCACCACCACCCTCACCCTCACCCACCCCCAC 900
              |||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db       9563 CTCACCCACCACAACCTCCCCAACACCCCCACCACCACCCTCACCCTCACCCACCCCCAC 9622

Qy        901 CACCACCCCCAAACCCACCTCCTC 924
              |||||||||||||||||||||||
Db       9623 CACCACCCCCAAACCCACCTCCTC 9646
```

RESULT 4
AF275943
LOCUS AF275943 11096 bp DNA linear BCT 02-SEP-2000
DEFINITION Streptomyces avermitilis avermectin polyketide synthase gene, partial cds.
ACCESSION AF275943
VERSION AF275943.1 GI:9964075
KEYWORDS .
SOURCE Streptomyces avermitilis.
ORGANISM Streptomyces avermitilis
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae; Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE 1 (bases 1 to 11096)
AUTHORS Hong,Y.-S. and Lee,J.J.
TITLE Targeted Gene Disruption of the avermectin O-methyltransferase gene and polyketide synthase gene from Streptomyces avermitilis
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 11096)
AUTHORS Hong,Y.-S. and Lee,J.J.
TITLE Direct Submission
JOURNAL Submitted (07-JUN-2000) Anticancer Agent Research Laboratory, Korea Research Institute of Bioscience and Biotechnology, P.O. Box 116, YuSong-Gu, Taejon 305-600, South Korea
FEATURES Location/Qualifiers
source 1. .11096
/organism="Streptomyces avermitilis"
/strain="ATCC31271"
/db_xref="ATCC:31271"
/db_xref="taxon:33903"
CDS 220. .>11096
/note="Avr1"
/codon_start=1
/transl_table=11
/product="avermectin polyketide synthase"
/protein_id="AAG09812.1"
/db_xref="GI:9964076"
/translation="MADEADGGVVFVFPGQGPQWPGMGRELLDASDVFRESVRACEAE
FAPYVDWSVEQVLRDSPDAPGLDRVDVVQRTLFAVMISLAALWRSQGVEPCAVLGHSL
GEIAAAHVSGGLSLADAARVGDAWSQAQTTLAGTGALVSVAATPDELLPRIAPWTEDN
PARLAVAAVNGPRSTVVSGAREAVADLVADLTAAQVRTRMIPVDVPAHSPLMYAIEER
VVSGLLPITPRPSRIPFHSSVTGGRLDTRELDAAYWYRNMSSTVRFEPAARLLLQQGP
KTFVEMSPHPVLTMGLQELAADLGDTTGTADTVIMGTLRRGQGTLDHFLTSLAQLRGH
GETSATTVLSARLTALSPTQQQSLLLDLVRAHTMAVLNDDGNERTASDAGPSASFAHL
GFDSVMGVELRNRLSKATGLRLPVTLILDHTTPAAVAARLRTAALGHLDEDTAPVPDS
PSGHGGTAAADDPIAIIGMACRFPGGVRSPKDLWELAAASGGDAIGPFPTDRGWPTEQR
HAQDPTQPGTFYPQGGGFLHDAAHFDAGFFGISPREALAMDPQQRLLLETSWEAFERA
GIDPLSVRGSRTGVFAGALSFDYGPRMDTASSEGAADVEGHILTGTTGSVLSGRSAYS
FGLEGPAITVDTGCSASLVTLHLACQSLRSGECTFALAGGVSDVHPGMFIEFSRQCGL
SVDGRCKAVSAAADGTGWGEGIGLLAVVRGSAVNQDGASNGLTAPNGPAQERVIRQAL
ANAGLSVADVDVVEGHGTGTTLGDPIEAQALLATYGQRAGDRPLWLGSLKSNIGHTMA
AAGVGGVIKMVMALREGVLPRTLHVDEPSPQGLAAGAVRLLTEAVPWPGDAAGRLRRA
GVSSFGIGGTNAHVILEEAPAAGGCVAGGRVLEGAPGLAISVAESVAAPVAVSAPVAE
SVPVPVPVPVPVPVSARSEAGLRAQAEALRQYEAVQPDVSLADVGAGLACRQAVLEHH
VVILAACTSSSRSAARTTARSSSTARPQARPAPTSARRIGALAAGSGSAALTTGHAPG
GDRGGVVFVFPWQGGQWAGMGVRLLCLLRVFARRMQACEEALAPWVDWSVVDILRRDA
GDAVWEQADVVQPVLFSVMVSLAALWRSYGIEPNEVLGHSKDEIAAAHIYGALSLKDA
AKTVALPPQEVEQLIGERGGRLWVAAVNGPRSTAVSGDAEAVVEVLAYCAGTGVRPRI
PVDYASHCPHVQPLREELLELLGDISPQPYGVPFFSTVEGTWLDTTTLDAAYWYRNLH
QPVRFSHDVQALADQGHRVLLEVSPHPTLVPAIEDTTEDTARRRHCDRQPPPRRERHP
LLPQRLRLDPYYRHRQTHHVAPLLDPPRHLPPPLDAPRPAHLPFQHQHYWLESSQPGA
GSGSGAGAGSGAGSGRAGTAGGTAEVESRFWDAVARQDLETVATTLAVPPSAGLDTVV
PALSAWHRHQHDQARINTWTYQETWKPLTLPTTHQPHQTWLIAIPETQTHHPHITNIL
TNLHHHGITPIPLPLTTHHTNPQHLHHTLHHTRQQAQNHTTGAITGLLSLLALDETPH
PHHPHTPTGTLLNLTLTQTHTQTHPPTPLWYATTNATTTHPNDPLTHPTQAQTWGLAR
TTLLEHPTHTAGIIDLPTTPTPHTLHHLTQTLTQPHHQTQLAIRTTGTHTRRLTPTTL
TPTHQPPTPTPHGTTLITGGTGALATHLTHHLTTHQPTQHLLLTSRTGPHTPHAQHLT
TQLQQKGIHLTITTCDTSTPRPTHNNSLNTIPPQHPVTTVIHTGGILDDATLTNLTPT
QLNNVLRAKAHSAHLLHQLTQHTPLTAFVLYSSAAATFGAPGQANYAAANAYLDALAH
HRHTHHLPATSIAWGTWQGNGLADSDKARAYLDRRGFRPMSPELATAAVTQAIADTER
PYVVIADIDWSKIEHTSQTSDLVSAARAREPAVQRPTPPAELHKTLAHQTSADQRAAL
LELVRDHVAAVLRHADPKAIAPDQSFRALGFDSLTAVEFRNLLIKATGLRLPVSLVFD
HPTPAKLAVHLQNQLRGTAAESAPSAAAVTAEASVTEPIAIVGMACRFPGGVTSADDF
WDLISSEQDAIGGFPTDRGWDLDTLYDPDPDHPGTCYTRNGGFLYDAGHFDAEFFGIS
PREALAMDPQQRLLLETAWETIEHAGINPHTLHGTPTGVFTGTNGQDYALRVHNAGQS
TDGFALTGTAGSVISGRISYTFGFEGPAVSVDTACSSSLVALHLACQALRAGECSMAL
AGGVTVMSSPGAFVEFSRQRGLAADGHCKAFSAAADGTGWGEGVGMLLVERLSDAHRN
GHRVLAVVRGSAVNQDGASNGLTAPNGPSQQRVIRQALANAGLSAGDVDAVEAHGTGT
TLGDPIEAQALLATYGQDRAGEGPLWLGSVKSNVGHTQAAAGVAGVIKMVMALRHGLL
PRTLHVDEPSPHVDWSAGAVQLLTETVPWPGGEGRLRRAGVSSFGVSGTNAHVILEEA
PADDVPGGPPAGEGDAGSDDEAAAGSPGVWPWLVSAKSQPALRAQAQALHAHLTDHPG
LDLADVGYTLAHARAVFDHRATLIAADRDTFLQALQALAAGEPHPAVIHSSAPGGTGT
GEAAGKTAFICSGQGTQRPGMAHGLYHTHPVFAAALNDICTHLDPHLDHPLLPLLTQN
DNDNEDAAALLQQTRYAQPALFAFQVALHRLLTDGYHITPHYYAGHSLGEITGAHLAG
ILTLTDATTLITQRATLMQTMPPGTMTTLHTTPHHITHHLTAHENDLAIAAINTPTSL
VISGTPHTVQHITTLCQQQGIKTKTLPTNHAFHSPHTNPILNQLHQHTQTLTYHPPHT
PLITANTPPDQLLTPHYWTQQARNTVDYATTTQTLHQHGVTTYIELGPDNTLTTLTHH
NLPNTPTTTLTLTHPHHHPQTHLLTNLAKTTTTWHPHHYTHHAQPTPHPHPLDLPTYP
FQHHHYWLESTQPGAGNVSAAGLDPTEHPLLGATLELATDGGALLAGRLSLRSHPWLA
DHAVGGTVLLSGATFLELALHAGTYVGCDRVDELTLHAPLVVPVDGGVSVQVGVAAAD
GKGRRLVSVYARGGSACGGGASGGVWTCHASGVLVEAAAGGVVVDGLAGVWPPRGAV
AVDVDGVRDRLAGAGCVLGPVFSGLRAVWRDGGDLLAEVCLPEEAWGDAAGFGLHPAL
LNGVVQPLSVLLPGGTGFGKGAGFGKGVRVPAVWGGVSLHRAGVTGVRVRVSAVGRGG
GREAVSVVVGDEAGVPVASVDRLELRPVDMGQLRAVSVSAGRRGSLYAVQWAEVGPVP
VCGQAWAWHEDVGESGGGPVPGVVVLRCPDAGAGGGGGGGGGGVGEVVGGVLGVVQG
WLGLERFAGSRLVVVTRGAVVAGPEDGPVDVVGASVWGLVRSAQAEHPDRFVLLDLDT
DTGTDLDTGAGAGWGVDGGRVAAVVACGEPQLAVRGERLLAARLTRLESSGDVPAQRS
GDTRARRSDVPAQRSGGVPARRSVDVSGREVLPWLSGGSVLVTGGTGVLGAAVARHLA
GVCGVRDLLLVSRRGPDAPGAEGLRRSWPRGAEVRIVACDVGERREVVRLAGGCSCRV
SVPVDFSAVASVTVRLPVASDVRKEAAMAYATVEEFTDYLDPDP"

```
BASE COUNT     1756 a   4059 c   3528 g   1753 t
ORIGIN

  Query Match           99.3%;  Score 917.6;  DB 1;  Length 11096;
  Best Local Similarity  99.6%;  Pred. No. 2.7e-88;
  Matches  920;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy          1 ACCGCATTCATCTGCTCCGGACAGGGCACCCAACGCCCCGGCATGGCCCACGGCCTCTAC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8026 ACCGCATTCATCTGCTCCGGACAGGGCACCCAACGCCCCGGCATGGCCCACGGCCTCTAC 8085

Qy         61 CACACCCACCCCGTCTTCGCCGCCGCACTCAACGACATCTGCACCCACCTCGACCCCCAC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8086 CACACCCACCCCGTCTTCGCCGCCGCACTCAACGACATCTGCACCCACCTCGACCCCCAC 8145

Qy        121 CTCGACCACCCCCTCCTCCCCCTCCTCACCCAAAACGACAACGACAACGAGGACGCGGCC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8146 CTCGACCACCCCCTCCTCCCCCTCCTCACCCAAAACGACAACGACAACGAGGACGCGGCC 8205

Qy        181 GCACTGCTCCAGCAGACCCGCTACGCCCAGCCCGCCCTCTTCGCCTTCCAGGTCGCCCTC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8206 GCACTGCTCCAGCAGACCCGCTACGCCCAGCCCGCCCTCTTCGCCTTCCAGGTCGCCCTC 8265

Qy        241 CACCGCCTCCTCACCGACGGCTACCACATCACCCCCCACTACTACGCCGGACACTCCCTC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8266 CACCGCCTCCTCACCGACGGCTACCACATCACCCCCCACTACTACGCCGGACACTCCCTC 8325

Qy        301 GGCGAAATCACCGCCGCCCACCTCGCCGGCATCCTCACCCTCACCGACGCCACCACCCTC 360
              |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db       8326 GGCGAAATCACCGGCGCCCACCTCGCCGGCATCCTCACCCTCACCGACGCCACCACCCTC 8385

Qy        361 ATCACCCAACGCGCCACCCTCATGCAAACCATGCCCCCCGGCACCATGACCACCCTCCAC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db       8386 ATCACCCAACGCGCCACCCTCATGCAAACCATGCCCCCCGGCACCATGACCACCCTGCAC 8445

Qy        421 ACCACCCCCCCACCACATCACCCACCACCTCACCGCCCACGAAAACGACCTCGCCATCGCC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8446 ACCACCCCCCCACCACATCACCCACCACCTCACCGCCCACGAAAACGACCTCGCCATCGCC 8505

Qy        481 GCCATCAACACCCCCACCTCCCTCGTCATCAGCGGCACCCCCCACACCGTCCAACACATC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8506 GCCATCAACACCCCCACCTCCCTCGTCATCAGCGGCACCCCCCACACCGTCCAACACATC 8565

Qy        541 ACCACCCTCTGCCAACAACAAGGCATCAAAACCAAAACCCTCCCCACCAACCACGCCTTC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8566 ACCACCCTCTGCCAACAACAAGGCATCAAAACCAAAACCCTCCCCACCAACCACGCCTTC 8625

Qy        601 CACTCCCCCCACACCAACCCCATCCTCAACCAACTCCACCAGCACACCCAAACCCTCACC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8626 CACTCCCCCCACACCAACCCCATCCTCAACCAACTCCACCAGCACACCCAAACCCTCACC 8685

Qy        661 TACCACCCACCCCACACCCCCCTCATCACCGCCAACACCCCACCCGACCAACTCCTCACC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8686 TACCACCCACCCCACACCCCCCTCATCACCGCCAACACCCCACCCGACCAACTCCTCACC 8745

Qy        721 CCCCACTACTGGACCCAACAAGCCCGCAACACCGTCGACTACGCCACCACCACCCAAACC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8746 CCCCACTACTGGACCCAACAAGCCCGCAACACCGTCGACTACGCCACCACCACCCAAACC 8805

Qy        781 CTCCACCAACACGGCGTCACCACCTACATCGAACTCGGACCCGACAACACCCTCACCACC 840
              || ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8806 CTTCACCAACACGGCGTCACCACCTACATCGAACTCGGACCCGACAACACCCTCACCACC 8865

Qy        841 CTCACCCACCACAACCTCCCCAACCCCCCCACCACCACCCTCACCCTCACCCACCCCCAC 900
              |||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db       8866 CTCACCCACCACAACCTCCCCAACACCCCCACCACCACCCTCACCCTCACCCACCCCCAC 8925

Qy        901 CACCACCCCCAAACCCACCTCCTC 924
              |||||||||||||||||||||||||
Db       8926 CACCACCCCCAAACCCACCTCCTC 8949


RESULT 5
E38021
LOCUS       E38021                31422 bp    DNA     linear   PAT 31-JAN-2002
DEFINITION  Avermectin aglycon synthase gene.
ACCESSION   E38021
VERSION     E38021.1  GI:18626910
KEYWORDS    JP 2000245457-A/2.
SOURCE      Streptomyces avermitilis.
  ORGANISM  Streptomyces avermitilis
            Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
            Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1  (bases 1 to 31422)
  AUTHORS   Omura,S. and Ikeda,H.
  TITLE     Avermectin aglycon synthase gene
  JOURNAL   Patent: JP 2000245457-A 2 12-SEP-2000;
            THE KITASATO INSTITUTE
COMMENT     OS   Streptomyces avermitilis
            PN   JP 2000245457-A/2
            PD   12-SEP-2000
            PF   24-FEB-1999 JP 1999046961
            PR
            PI   SATOSHI OMURA,HARUO IKEDA
            PC   C12N15/00,A61K31/70,C12N1/15,C12N9/88,C12P19/62, PC
            C12Q1/68//C07D493/22,
            PC   (C12N1/15,C12R1:465),(C12N9/88,C12R1:465),C12N15/00 CC
            FH   Key             Location/Qualifiers
            FT   CDS             (1). .(14643)
            FT   CDS             (14824). .(31419).
FEATURES             Location/Qualifiers
     source          1. .31422
                     /organism="Streptomyces avermitilis"
                     /db_xref="taxon:33903"
BASE COUNT     4136 a  10237 c  11677 g   5372 t
ORIGIN

  Query Match             94.5%;  Score 873.4;  DB 6;  Length 31422;
  Best Local Similarity   96.7%;  Pred. No. 9.7e-84;
  Matches  908;  Conservative    0;  Mismatches   16;  Indels   15;  Gaps    1;

Qy          1 ACCGCATTCATCTGCTCCGGACAGGGCACCCAACGCCCCGGCATGGCCCACGGCCTCTAC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      22090 ACCGCATTCATCTGCTCCGGACAGGGCACCCAACGCCCCGGCATGGCCCACGGCCTCTAC 22149

Qy         61 CACACCCACCCCGTCTTCGCCGCCGCACTCAACGACATCTGCACCCACCTCGACCCCCAC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      22150 CACACCCACCCCGTCTTCGCCGCCGCACTCAACGACATCTGCACCCACCTCGACCCCCAC 22209

Qy        121 CTCGACCACCCCCTCCTCCCCCTCCTCACCCAAAACGACAACGACAA------------- 167
              |||||||||||||||||||||||||||||||  ||  ||||  | |
Db      22210 CTCGACCACCCCCTCCTCCCCCTCCTCACCCAGGACCCCAACACCCAGGACACCACCACC 22269

Qy        168 --CGAGGACGCGGCCGCACTGCTCCAGCAGACCCGCTACGCCCAGCCCGCCCTCTTCGCC 225
                ||| || |||||||||||||||||||||||||  ||||||||||||||||||||||||
Db      22270 CTCGAAGAAGCGGCCGCACTGCTCCAGCAGACCCCGTACGCCCAGCCCGCCCTCTTCGCC 22329

Qy        226 TTCCAGGTCGCCCTCCACCGCCTCCTCACCGACGGCTACCACATCACCCCCCACTACTAC 285
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      22330 TTCCAGGTCGCCCTCCACCGCCTCCTCACCGACGGCTACCACATCACCCCCCACTACTAC 22389

Qy        286 GCCGGACACTCCCTCGGCGAAATCACCGCCGCCCACCTCGCCGGCATCCTCACCCTCACC 345
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      22390 GCCGGACACTCCCTCGGCGAAATCACCGCCGCCCACCTCGCCGGCATCCTCACCCTCACC 22449

Qy        346 GACGCCACCACCCTCATCACCCAACGCGCCACCCTCATGCAAACCATGCCCCCCGGCACC 405
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      22450 GACGCCACCACCCTCATCACCCAACGCGCCACCCTCATGCAAACCATGCCCCCCGGCACC 22509

Qy        406 ATGACCACCCTCCACACCACCCCCCCACCACATCACCCACCACCTCACCGCCCACGAAAAC 465
              ||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db      22510 ATGACCACCCTCCACACCACCCCCCACCACATCACCCACCACATCACCGCCCACGAAAAC 22569

Qy        466 GACCTCGCCATCGCCGCCATCAACACCCCCACCTCCCTCGTCATCAGCGGCACCCCCCAC 525
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      22570 GACCTCGCCATCGCCGCCATCAACACCCCCACCTCCCTCGTCATCAGCGGCACCCCCCAC 22629

Qy        526 ACCGTCCAACACATCACCACCCTCTGCCAACAACAAGGCATCAAAACCAAAACCCTCCCC 585
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      22630 ACCGTCCAACACATCACCACCCTCTGCCAACAACAAGGCATCAAAACCAAAACCCTCCCC 22689

Qy        586 ACCAACCACGCCTTCCACTCCCCCCACACCAACCCCATCCTCAACCAACTCCACCAGCAC 645
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      22690 ACCAACCACGCCTTCCACTCCCCCCACACCAACCCCATCCTCAACCAACTCCACCAGCAC 22749

Qy        646 ACCCAAACCCTCACCTACCACCCACCCCACACCCCCCTCATCACCGCCAACACCCCACCC 705
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      22750 ACCCAAACCCTCACCTACCACCCACCCCACACCCCCCTCATCACCGCCAACACCCCACCC 22809

Qy        706 GACCAACTCCTCACCCCCCACTACTGGACCCAACAAGCCCGCAACACCGTCGACTACGCC 765
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||
Db      22810 GACCAACTCCTCACCCCCCACTACTGGACCCAACAAGCCCGCAACACCGTCGACATAGCC 22869

Qy        766 ACCACCACCCAAACCCTCCACCAACACGGCGTCACCACCTACATCGAACTCGGACCCGAC 825
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      22870 ACCACCACCCAAACCCTCCACCAACACGGCGTCACCACCTACATCGAACTCGGACCCGAC 22929

Qy        826 AACACCCTCACCACCCTCACCCACCACAACCTCCCCAACCCCCCCACCACCACCCTCACC 885
              ||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db      22930 AACACCCTCACCACCCTCACCCACCACAACCTCCCCAACACCCCCACCACCACCCTCACC 22989

Qy        886 CTCACCCACCCCCACCACCACCCCCAAACCCACCTCCTC 924
              |||||||||||||||||||||||||||||||||||||||
Db      22990 CTCACCCACCCCCACCACCACCCCCAAACCCACCTCCTC 23028
```

```
RESULT 6
AB032367/c
LOCUS       AB032367               64957 bp    DNA     linear   BCT 18-SEP-1999
DEFINITION  Streptomyces avermitilis polyketide synthase gene cluster (aveA1,
            aveA2, aveA3, aveA4) and aveC, aveE genes, complete cds.
ACCESSION   AB032367
VERSION     AB032367.1  GI:5902890
KEYWORDS    AveC; cytochrome P450 hydroxylase; type I polyketide synthase AVES
            4; type I polyketide synthase AVES 3; type I polyketide synthase
            AVES 2; type I polyketide synthase AVES 1.
SOURCE      Streptomyces avermitilis DNA.
  ORGANISM  Streptomyces avermitilis
            Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
            Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1  (bases 1 to 64957)
  AUTHORS   Ikeda,H., Nonomiya,T., Usami,M., Ohta,T. and Omura,S.
  TITLE     Organization of the biosynthetic gene cluster for the polyketide
            anthelmintic macrolide avermectin in Streptomyces avermitilis
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9509-9514 (1999)
  MEDLINE   99380548
REFERENCE   2  (bases 1 to 64957)
  AUTHORS   Ikeda,H., Nonomiya,T., Usami,M., Ohta,T. and Omura,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-SEP-1999) Haruo Ikeda, School of Pharmaceutical
            Sciences, Kitasato University, Microbial chemistry; 5-9-1
            Shirokane, Minato-ku, Tokyo 108-8641, Japan
            (E-mail:ikeda@mc.pharm.kitasato-u.ac.jp, Tel:+81-3-5791-6242,
            Fax:+81-3-3444-6197)
FEATURES             Location/Qualifiers
     source          1. .64957
                     /organism="Streptomyces avermitilis"
                     /db_xref="taxon:33903"
     gene            101. .12019
                     /gene="aveA1"
     CDS             101. .12019
                     /gene="aveA1"
                     /note="contains a loading module, Ave module 1 and module
                     2
                     multifunctional polyketide synthase"
                     /codon_start=1
                     /transl_table=11
                     /product="type I polyketide synthase AVES 1"
                     /protein_id="BAA84474.1"
                     /db_xref="GI:5902891"
                     /translation="MQRMDGGEEPRPAAGEVLGVADEADGGVVFVFPGQGPQWPGMGR
                     ELLDASDVFRESVRACEAAFAPYVDWSVEQVLRDSPDAPGLDRVDVVQPTLFAVMISL
                     AALWRSQGVEPCAVLGHSLGEIAAAHVSGGLSLADAARVVTLWSQAQTTLAGTGALVS
                     VAATPDELLPRIAPWTEDNPARLAVAAVNGPRSTVVSGAREAVADLVADLTAAQVRTR
                     MIPVDVPAHSPLMYAIEERVVSGLLPITPRPSRIPFHSSVTGGRLDTRELDAAYWYRN
                     MSSTVRFEPAARLLLQQGPKTFVEMSPHPVLTMGLQELAPDLGDTTGTADTVIMGTLR
                     RGQGTLDHFLTSLAQLRGHGETSATTVLSARLTALSPTQQQSLLLDLVRAHTMAVLND
                     DGNERTASDAGPSASFAHLGFDSVMGVELRNRLSKATGLRLPVTLIFDHTTPAAVAAR
                     LRTAALGHLDEDTAPVPDSPSGHGGTAAADDPIAIIGMACRFPGGVRSPKDLWELAAS
                     GGDAIGPFPTDRGWPTEQRHAQDPTQPGTFYPQGGGFLHDAAHFDAGFFGISPREALA
                     MDPQQRLLLETSWEAFERAGIDPLSVRGSRTGVFAGALSFDYGPRMDTASSEGAADVE
                     GHILTGTTGSVLSGRIAYSFGLEGPAITVDTGCSASLVTLHLACQSLRSGECTLALAG
                     GVSVMSTLGMFIEFSRQRGLSVDGRCKAYSAAADGTGWGEGVGMLLVERLSDAVRLGH
                     RVLAVVRGSAVNQDGASNGLTAPNGPAQERVIRQALANAGLSVADVDVVEGHGTGTTL
                     GDPIEAQALLATYGQRAGDRPLWLGSLKSNIGHTMAAAGVGGVIKMVMALREGVLPRT
                     LHVDKPSPQVDWSAGAVRLLTEAVPWPGDAAGRLRRAGVSSFGIGGTNAHVILEEAPA
                     AGGCVAGGGVLEGAPGLAISVAESVAAPVAVSAPVAESVPVPVPVPVPVSARSEAG
                     LRAQAEALRQYVAVRPDVSLADVGAGLACGRAVLEHRAVVLAADREELVQGLGALAAG
                     EPDRRVTTGHAPGGDRGGVVFVFPGQGGQWAGMGVRLLASSPVFARRMQACEEALAPW
                     VDWSVVDILRRDAGDAVWERADVVQPVLFSVMVSLAALWRSYGIEPDAVLGHSQGEIA
                     AAHVCGALSLKDAAKTVALRSRALAAVRGRGGMASVPLPAQEVEQLIGERWAGRLWVA
                     AVNGPRSTAVSGDAEAVDEVLAYCAGTGVRARRIPVDYASHCPHVQPLREELLELLGD
                     ISPQPSGVPFFSTVEGTWLDTTTLDAAYWYRNLHQPVRFSDAVQALADDGHRVFVEVS
                     PHPTLVPAIEDTTEDTAEDVTAIGSLRRGDNDTRRFLTALAHTHTTGIGTPTTWHHHY
                     THHHTHPHPHTHLDLPTYPFQHQHYWLESSQPGAGSGSGAGAGSGAGSGRAGTAGGTA
                     EVESRFWDAVARQDLETVATTLAVPPSAGLDTVVPALSAWHRHQHDQARINTWTYQET
                     WKPLTLPTTHQPHQTWLIAIPETQTHHPHITNILTNLHHHGITPIPLTLNHTHTNPQH
                     LHHTLHHTRQQAQNHTTGAITGLLSLLALDETPHPHHPHTPTGTLLNLTLTQTHTQTH
                     PPTPLWYATTNATTTHPNDPLTHPTQAQTWGLARTTLLEHPTHTAGIIDLPTTPTPHT
                     LQHLTQTLTQPHHQTQLAIRTTGTHTRRLTPTTLTPTHQPPTPTPHGTTLITGGTGAL
                     ATHLTHHLTTHQPTQHLLLTSRTGPHTPHAQHLTTQLQQKGIHLTITTCDTSNPDQLQ
                     QLLNTIPPQHPLTTVIHTAGILDDATLTNLTPTQLNNVLRAKAHSAHLLHQLTQHTPL
                     TAFVLYSSAAATFGAPGQANYAAANAYLDALAHHRHTHHLPATSIAWGTWQGNGLADS
                     DKARAYLDRRGFRPMSPELATAAVTQAIADTERPYVVIADIDWSKIEHTSQTSDLVSA
                     ARELREPAVQRPTPPAELHKTLAHQTSADQRAALLELVRDHVAAVLRHADPKAIAPDQS
                     FRALGFDSLTAVEFRNLLIKATGLRLPVSLVFDHPTPAKLAVHLQNQLRGTAAESAPS
                     AAAVTAEASVTEPIAIVGMACRFPGGVTSADDFWDLISSEQDAIGGFPTDRGWDLDTL
                     YDPDPDHPGTCYTRNGGFLYDAGHFDAEFFGISPREALAMDPQQRLLLETAWETIEHA
                     GINPHTLHGTPTGVFTGTNGQDYALRVHNAGQSTDGFALTGTAGSVISGRISYTFGFE
                     GPAVSVDTACSSSLVALHLACQALRAGECSMALAGGVTVMSSPGAFVEFSRQRGLAAD
                     GHCKAFSAAADGTGWGEGVGMLLVERLSDAHRNGHRVLAVVRGSAVNQDGASNGLTAP
                     NGPSQQRVIRQALANAGLSAGDVDAVEAHGTGTTLGDPIEAQALLATYGQDRAGEGPL
                     WLGSVKSNVGHTQAAAGVAGVIKMVMALRHGLLPRTLHVDEPSPHVDWSAGAVQLLTE
                     TVPWPGGEGRLRRAGVSSFGVSGTNAHVILEEAPADDVPGGPPAGEGDAGSDDEAAAG
                     SPGVWPWLVSAKSQPALRAQAQALHAHLTDHPGLDLADVGYTLAHARAVFDHRATLIA
                     ADRDTFLQALQALAAGEPHPAVIHSSAPGGTGTGEAAGKTAFICSGQGTQRPGMAHGL
                     YHTHPVFAAALNDICTHLDPHLDHPLLPLLTQNDNDNEDAAALLQQTRYAQPALFAFQ
                     VALHRLLTDGYHITPHYYAGHSLGEITAAHLAGILTLTDATTLITQRATLMQTMPPGT
                     MTTLHTTPHHITHHLTAHENDLAIAAINTPTSLVISGTPHTVQHITTLCQQQGIKTKT
                     LPTNHAFHSPHTNPILNQLHQHTQTLTYHPPHTPLITANTPPDQLLTPHYWTQQARNT
                     VDYATTTQTLHQHGVTTYIELGPDNTLTTLTHHNLPNPPTTTLTLTHPHHHPQTHLLT
                     NLAKTTTTWHPHHYTHHDNQPHTHTHLDLPTYPFQHHHYWLESTQPGAGNVSAAGLDP
                     TEHPLLGATLELATDGGALLAGRLSLRSHPWLADHAVGGTVLLSGATFLELALHAGTY
                     VGCDRVDELTLHAPLVVPVDGGVSVQVGVAAADGEGRRLVSVYARGGSACGGGGASGG
                     VWTCHASGVLVEAAAGGVVVDGLAGVWPPRGAVAVDVDGVRDRLAGAGCVLGPVFSGL
                     RAVWRDGGDLLAEVCLPEEAWGDAAGFGLHPALLDGVVQPLSVLLPGGTGFGEGAGFG
                     EGVRVPAVWGGVSLHRAGVTGVRVRVSAVGRGGGREAVSVVVGDEAGVPVASVDRLEL
                     RPVDMGQLRAVSVSAGRRGSLYAVQWAEVGPVPVCGQAWAWHEDVGESGGGPVPGVVV
                     LRCPDAGAGGGGGGGGGGVGEVVGGVLGVVQGWLGLERFAGSRLVVVTRGAVVAGPE
                     DGPVDVVGASVWGLVRSAQAEHPDRFVLLDLDTDTGTDLDTGAGAGWGVDGGRVAAVV
                     ACGEPQLAVRGERLLAARLKRLESSGDVPAQRSGDTRARRSDVPAQRSGGVPARRSVD
                     VSGREVLPWLSGGSVLVTGGTGVLGAAVARHLAGVCGVRDLLLVSRRGPDAPGAEGLR
                     AELAALGAEVRIVACDVGERREVVRLLEGVPAGCPLTGVVHAAGVLDDATIASLTPER
                     LGTVFAAKVDAALLLDELTRGMELSAFVLFSSAAGILGSAGQGNYAAANAALDALAYR
                     RRAAGLPGVSLAWGLWEEASGMTGHLAGTDHRRIIRSGLHPMSTPDALALFDAALALD
                     RPVLLPADLRPAPPLPPLLQDLLPATRRRTTRTTTTGGADNGAQLHARLAGQTHEQQH
                     TTLLALVRSHIATVLGHTTPDTIPPDRAFRDLGFDSLTAVELRNRLSRTTGLRLPTTL
                     AFDHPNPTTLTHHLHTQLQPQPDNAVAPVLAELDKLESALSALDKTDSASERVTLRLK
                     SLMLRWNAPQHPTAESADDDEKFTSATEAEIFKFIDNDLGLS"
gene                 12071. .30790
                     /gene="aveA2"
CDS                  12071. .30790
                     /gene="aveA2"
                     /note="contains Ave module 3, module 4, module 5 and
                     module 6
                     multifunctional polyketide synthase"
                     /codon_start=1
                     /transl_table=11
                     /product="type I polyketide synthase AVES 2"
                     /protein_id="BAA84475.1"
                     /db_xref="GI:5902892"
                     /translation="MQLANEAKLLEYLKRVTADLDRTRRRLYEVVEREQEPIAIVGMA
                     CRYPGGATSPTRLWHLVKSQTDAIGEFPTDRGWNLEQLYDPDPDRSGTSYTRSGGFLY
                     DAGDFDAAFFELSPREALAMDPQQRLLLETTWETFEQGGIDPRSMRGSRTGVFVGINP
                     EDYTTGYTHQPSNAVEGYLLTGSAASIASGRISYNFGLEGPAITIDTACSSSLVALHL
                     ACQALRSGECTMALAGGASVMATPFVFTEFSRQRGLAADGRCKAFSAAADGTGWSEGV
                     GMLLVERLSDARRNGHRVLAVVRGSAVNQDGASNGLTAPNGRSQVKVIRQALANAHLS
                     PADVDAVEAHGTGTTLGDPIEAQALVEAYGQDRPNGRPLWLGTLKSNIGHSMAAAGVG
                     GVIKMVMALRNGLLPRTLHVDEPSPHVDWSAGAVQLLTETVPWPGGEGRLRRAGVSSF
                     GVSGTNAHVILEEAPAHNIPSDTPADDVPGESAADEDAGSGDEAAAGSPGVWPWLVSA
                     KSQPALRAQAQALHAHLTDHPGLDLADVGYTLAHARAVFDHRATLIAADRDTFLQALQ
                     ALAAGEPHPAVIHSSAPGGTGTGEAAGKTAFICSGQGTQRPGMAHGLYHTHPVFAAAL
                     NDICTHLDPHLDHPLLPLLTQDPNTQDTTTLEEAAALLQQTRYAQPALFAFQVALHRL
                     LTDGYHITPHYYAGHSLGEITAAHLAGILTLTDATTLITQRATLMQTMPPGTMTTLHT
                     TPHHITHHLTAHENDLAIAAINTPTSLVISGTPHTVQHITTLCQQQGIKTKTLPTNHA
                     FHSPHTNPILNQLHQHTQTLTYHPPHTPLITANTPPDQLLTPHYWTQQARNTVDYATT
                     TQTLHQHGVTTYIELGPDNTLTTLTHDNLPNTPTTTLTLTHPHHHPQTHLLTNLAKTT
                     TTWHPHHYTHHHNQPHTHTHLDLPTYPFQHHHYWLQPPGKPSDPSPSEGREQATTPST
                     PLRDVLVGKSPQERDEELLRLVRTHAAAVLGHATPEVIVPNKAFKELGFDSLAAIQLR
                     NRLLADVDLPLPATLIFDYPTPMALCQFLRAAIVGADTGTTTRLPLTAVPADEPIAIV
                     GMACRYPGDVRTVDDLWQVVSGGHDAIGGFPTNRGWDLDTLYNPDPDHHGTSYTRSGG
```

```
FLYDAGNFDPDFFGISPREALAMDPQQRLLLETAWESIEHACINPDSLRGTPTGVFAG
LTYHDYAARFPTAPAGFEGYLGHGSAGSIASGRVAYALGLEGPALTVDTACSSSLVAL
HLACQALRSGECSMALAGGVTVMSTPAGFVEFSRQRGLAVDGRCKAFSAAADGTGWGE
GVGMLLVERLSDARRLGHRILAVVRGSAVNQDGASNGLTAPNGPSQERVIRLALANAD
LTPADVDAVEAHGTGTTLGDPIEAQALLATYGQDRPGNEPLWLGSMKSNIGHAQAAAG
VGGVIKMVMALRNGLLPRTLHVDEPSPHVDWSAGAVQLLTETVPWPGGEGRLRRAGVS
SFGVSGTNAHVILEEAPAHNIPSDTPADDAPGEAAADDVPGEAAGDDAGTGGEATGPA
AGSPGVWPWLVSAKSQPALRAQAQALHAHLTDHPGLDLADVGYTLAHARAVFDHRATL
IAADRDTFLQALQALAAGEPHPAVIHSSAPGGTGTGEAAGKTAFICSGQGTQRPGMAH
GLYHTHPVFAAALNDICTHLDPHLDHPLLPLLTQDPNTQDTTTLEEAAALLQQTPYAQ
PALFAFQVALHRLLTDGYHITPHYYAGHSLGEITAAHLAGILTLTDATTLITQRATLM
QTMPPGTMTTLHTTPHHITHHLTAHENDLAIAAINTPTSLVISGTPHTVQHITTLCQQ
QGIKTKTLPTKNAFHSPHTNPILNQLHQHTQTLTYHPPHTPLITANTPPDQLLTPHYW
TQQARNTVDYATTTQTLHQHGVTTYIELGPDNTLTTLTHHNLPNTPTTTLTLTHPHHH
PQTHLLTNLAKTTTTWHPHHYTHHHNQPHTHTHLDLPTYPFQHQHYWLESTQPGAGSG
SGSGSGRAGTAGGTAEVESRFWDAVARQDLETVATTLAVPPSAGLDTVVPALSAWHRH
QHDQARINTWTYQETWKPLTLPTTHQPHQTWLIAIPETQTHHPHITNILTNLHHHGIT
PIPLTLNHTHTNPQHLHHTRQQAQNHTTGPITGLLSLLALDETPHPHHPHTPTGTLLN
LTLTQTHTQTHPPTPLWYATTNATTTHPNDPLTHPTQAQTWGLARTTLLEHPTHTAGI
IDLPTTPTPHTLHHLTQTLTQPHHQTQLAIRTTGTHTRRLTPTTLTPTHQPPTPTPHG
TTLITGGTGALATHLTHHLTTHQPTQHLLLTSRTGPHTPHAQHLTTQLQQKGIHLTIT
TCDTSNPDQLQQLLNTIPPQHPLTTVIHTAGILDDATLTNLTPTQLNNVLRAKAHSAH
LLHQLTQHTPLNAFVLYSSAAATFGAPGQANYAAANAYLDALAHHRHTHHLPATSIAW
GTWQGNGLATGQVSEHLRRRGMFAMPPELAVTAVDGAIASGRPSLLVADIDWKKLGPV
LSSKSSVLLEDLPQAQGTEEARSTVEQTESTNLRQLLMGRSRSEQEEELLSLVRIHSA
AVLGRDDSEAIPPGRLFRDLGFDSLAAVELRNHLAAQTELALPTTLVFDYPSPTKLAQ
FLLSEIAEFQPDNSTPLPRPRAELDEPIAIVGMACRFPGGVTSADDFWDLISSEQDAI
GGFPTDRGWDLDTLYDPDPDHPGTCYTRNGGFLYDAGHFDAEFFGISPREALAMDPQQ
RLLLETAWETIEHAGINPHTLHGTPTGVFTGTNGQDHAAHIRQAPSGTEGFVLTGAAT
SIASGRISYILGLEGPAVTLDTACSSSLVALHLACQSLRSGECTMALAGGATVMTTPI
TFTEFARQRGLAPDGRCKAFSAAADGTGWGEGVGMLLVERLSDARRNGHRVLAVVRGS
AVNQDGASNGLTAPNGPSQQRVIRQALANADLTPADVDAVEAHGTGTTLGDPIEAQAI
LATYGQDRPGNGPLWLGSVKSNVGHTQAAAGVAGVIKMVMALRHRTLPPTLHADEPSP
HVDWSAGAVQLLTETVPWPGGEGRPRRAGVSSFGVSGTNAHVILEEAPADDVPGGPPA
DEDAGSGEEAAAGSPGVWPWLVSAKSQPALRAQAQALHAHLTDHPGLDLADVGYTLAH
ARAVFDHRATLIAADRDTFLQALQALAAGEPHPAVIHSSAPGGTGTGEAAGKTAFICS
GQGTQRPGMAHGLYHTHPVFAAALNDICTHLDPHLDHPLLPLLTQNDNDNDNEDAAAL
LQQTPYAQPALFAFQVALHRLLTDGYHITPHYYAGHSLGEITAAHLAGILTLTDATTL
ITQRATLMQTMPPGTMTTLHTTPHHITHHLTAHENDLAIAAINTPTSLVISGTPHTVQ
HITTLCQQQGIKTKTLPTNHAFHSPHTNPILNQLHQHTQTLTYHPPHTPLITANTPPD
QLLTPHYWTQQARNTVDYATTTQTLHQHGVTTYIELGPDNTLTTLTHHNLPNTPTTTL
TLTHPHHHPQTHLLTNLAKTTTTWHPHHYTHHHNQPHTHTHLDLPTYPFQHHHYWLEL
PSAQTSPGQRRSRRSAPDTAESEFWDAVNEEDLQSLAETLDIDASALDTVVPALSAWH
RHQHDQARINTWTYQETWKPLTLPTTHQPHQTWLIAIPETQTHHPHITNILTNLHHHG
ITPIPLTVNHTHTNPQHLHHTLHHTRQQAQNHTTGPITGLLSLLALDETPHPHHPHTP
TGTLLNLTLPQTHTQTHPPTPLWYATTNATTTHPNDPLTHPTQAQTWGLARTTLLEHP
THTAGIIDLPTTPTPHTLHHLTQTLTQPHHQTQLAIRTTGTHTRRLTPTTLTPTHQPP
TPTPHGTTLITGGTGALATHLTHHLTTHQPTQHLLLTSRTGPHTPHAQHLTTQLQQKG
IHLTITTCDTSNPDQLQQLLNTIPPQHPLTTVIHTAGVNLFAPVSETDAESFSSVTAA
KATGAAILHELLLDHETLEHFILFSSGAGAWGSGNQCAYSAANAYLDALATHRQTHGL
PGASIAWGPWAGKGMSAGDAAHGYLEKRGILPMEPRMALAAFHRARAQRPNSNLIIAD
IDWERFVPAFTARRHSPLIEDIPEVRQAAQELEAAASTAKTTTAQPIATSLRERLARL
TSSKQNQVLLGLIRTGICTVLGLRNPEGIEDQRAFRDLGFDSLTSAQFSKELAKETGL
PLPPSLVFDYPTPQECAAHLRTQLVDLDDEEDAALSNALPQVAHRRTVEDEPIAIIGM
ACRFPGGVRSADDLWELLASGKDAIGVFPTDRGWDLDTLYDPDPDHPGTCYTRNGGFL
YGAGHFDAEFFGISPREALAMDPQQRLLLETAWETIEHAGINPHTLHGTPTGVFAGIN
AQDHAAHIRQSRDVETIEGYALTGSSGSVASGRVAYTLGLEGPAVSVDTACSSSLVAL
```

```
Query Match          94.5%;  Score 873.4;  DB 1;  Length 64957;
Best Local Similarity  96.7%;  Pred. No. 8.4e-84;
Matches  908;  Conservative    0;  Mismatches   16;  Indels   15;  Gaps    1;

Qy          1 ACCGCATTCATCTGCTCCGGACAGGGCACCCAACGCCCCGGCATGGCCCACGGCCTCTAC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      42768 ACCGCATTCATCTGCTCCGGACAGGGCACCCAACGCCCCGGCATGGCCCACGGCCTCTAC 42709

Qy         61 CACACCCACCCCGTCTTCGCCGCCGCACTCAACGACATCTGCACCCACCTCGACCCCCAC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      42708 CACACCCACCCCGTCTTCGCCGCCGCACTCAACGACATCTGCACCCACCTCGACCCCCAC 42649

Qy        121 CTCGACCACCCCCTCCTCCCCCTCCTCACCCAAAACGACAACGACAA---------- 167
              |||||||||||||||||||||||||||||||  ||  ||||  | |
Db      42648 CTCGACCACCCCCTCCTCCCCCTCCTCACCCAGGACCCCAACACCCAGGACACCACCACC 42589
```

```
Qy        168 --CGAGGACGCGGCCGCACTGCTCCAGCAGACCCGCTACGCCCAGCCCGCCCTCTTCGCC 225
               ||| || ||||||||||||||||||||||||||  ||||||||||||||||||||||||
Db      42588 CTCGAAGAAGCGGCCGCACTGCTCCAGCAGACCCCGTACGCCCAGCCCGCCCTCTTCGCC 42529

Qy        226 TTCCAGGTCGCCCTCCACCGCCTCCTCACCGACGGCTACCACATCACCCCCCACTACTAC 285
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      42528 TTCCAGGTCGCCCTCCACCGCCTCCTCACCGACGGCTACCACATCACCCCCCACTACTAC 42469

Qy        286 GCCGGACACTCCCTCGGCGAAATCACCGCCGCCCACCTCGCCGGCATCCTCACCCTCACC 345
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      42468 GCCGGACACTCCCTCGGCGAAATCACCGCCGCCCACCTCGCCGGCATCCTCACCCTCACC 42409

Qy        346 GACGCCACCACCCTCATCACCCAACGCGCCACCCTCATGCAAACCATGCCCCCCGGCACC 405
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      42408 GACGCCACCACCCTCATCACCCAACGCGCCACCCTCATGCAAACCATGCCCCCCGGCACC 42349

Qy        406 ATGACCACCCTCCACACCACCCCCCCACCACATCACCCACCACCTCACCGCCCACGAAAAC 465
              |||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db      42348 ATGACCACCCTCCACACCACCCCCCACCACATCACCCACCACATCACCGCCCACGAAAAC 42289

Qy        466 GACCTCGCCATCGCCGCCATCAACACCCCCACCTCCCTCGTCATCAGCGGCACCCCCCAC 525
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      42288 GACCTCGCCATCGCCGCCATCAACACCCCCACCTCCCTCGTCATCAGCGGCACCCCCCAC 42229

Qy        526 ACCGTCCAACACATCACCACCCTCTGCCAACAACAAGGCATCAAAACCAAAACCCTCCCC 585
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      42228 ACCGTCCAACACATCACCACCCTCTGCCAACAACAAGGCATCAAAACCAAAACCCTCCCC 42169

Qy        586 ACCAACCACGCCTTCCACTCCCCCCACACCAACCCCATCCTCAACCAACTCCACCAGCAC 645
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      42168 ACCAACCACGCCTTCCACTCCCCCCACACCAACCCCATCCTCAACCAACTCCACCAGCAC 42109

Qy        646 ACCCAAACCCTCACCTACCACCCACCCCACACCCCCCTCATCACCGCCAACACCCCACCC 705
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      42108 ACCCAAACCCTCACCTACCACCCACCCCACACCCCCCTCATCACCGCCAACACCCCACCC 42049

Qy        706 GACCAACTCCTCACCCCCCACTACTGGACCCAACAAGCCCGCAACACCGTCGACTACGCC 765
              |||||||||||||||||||||||||||||||||||||||||||||||||||   |||
Db      42048 GACCAACTCCTCACCCCCCACTACTGGACCCAACAAGCCCGCAACACCGTCGACATAGCC 41989

Qy        766 ACCACCACCCAAACCCTCCACCAACACGGCGTCACCACCTACATCGAACTCGGACCCGAC 825
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      41988 ACCACCACCCAAACCCTCCACCAACACGGCGTCACCACCTACATCGAACTCGGACCCGAC 41929

Qy        826 AACACCCTCACCACCCTCACCCACCACAACCTCCCCAACCCCCCCACCACCACCCTCACC 885
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      41928 AACACCCTCACCACCCTCACCCACCACAACCTCCCCAACACCCCCACCACCACCCTCACC 41869

Qy        886 CTCACCCACCCCCACCACCACCCCCAAACCCACCTCCTC 924
              |||||||||||||||||||||||||||||||||||||||
Db      41868 CTCACCCACCCCCACCACCACCCCCAAACCCACCTCCTC 41830
```

```
RESULT 7
SC2C4/c
LOCUS       SC2C4                  24225 bp    DNA     linear   BCT 12-MAY-2002
DEFINITION  Streptomyces coelicolor cosmid 2C4.
ACCESSION   AL512902 AL645882
VERSION     AL512902.2  GI:20520683
KEYWORDS    accA1; secreted FAD-binding protein; type I polyketide synthase.
SOURCE      Streptomyces coelicolor A3(2).
  ORGANISM  Streptomyces coelicolor A3(2)
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1  (bases 1 to 24225)
  AUTHORS   Redenbach,M., Kieser,H.M., Denapaite,D., Eichner,A., Cullum,J.,
            Kinashi,H. and Hopwood,D.A.
  TITLE     A set of ordered cosmids and a detailed genetic and physical map
            for the 8 Mb Streptomyces coelicolor A3(2) chromosome
  JOURNAL   Mol. Microbiol. 21 (1), 77-96 (1996)
  MEDLINE   97000351
   PUBMED   8843436
```

```
REFERENCE   2  (bases 1 to 24225)
  AUTHORS   Seeger,K.J. and Harris,D.
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 24225)
  AUTHORS   Cerdeno,A.M., Parkhill,J., Barrell,B.G. and Rajandream,M.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-JAN-2001) Streptomyces coelicolor sequencing project,
            Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge
            CB10 1SA E-mail: barrell@sanger.ac.uk Cosmids supplied by Prof.
            David A. Hopwood, [3] John Innes Centre, Norwich Research Park,
            Colney, Norwich, Norfolk NR4 7UH, UK
COMMENT     On May 9, 2002 this sequence version replaced gi:12274798.
            Notes:
            Streptomyces coelicolor sequencing at The Sanger Centre is funded
            by the BBSRC and Beowulf Genomics
            Details of S. coelicolor sequencing at the Sanger Centre are
            available on the World Wide Web.
            (URL; http://www.sanger.ac.uk/Projects/S_coelicolor/) CDS are
            numbered using the following system eg SC7B7.01c. SC (S.
            coelicolor), 7B7 (cosmid name), .01 (first CDS), c (complementary
            strand).
            The more significant matches with motifs in the PROSITE database
            are also included but some of these may be fortuitous. The length
            in codons is given for each CDS.
            Usually the highest scoring match found by fasta -o is given for
            CDS which show significant similarity to other CDS in the database.
            The position of possible ribosome binding site sequences are given
            where these have been used to deduce the initiation codon. Gene
            prediction is based on positional base preference in codons using a
            specially developed Hidden Markov Model (Krogh et al., Nucleic
            Acids Research, 22(22):4768-4778(1994)) and the FramePlot program
            of Bibb et al., Gene 30:157-66(1984) as implemented at
            http://www.nih.go.jp/
            jun/cgi-bin/frameplot.pl. CAUTION:  We may not have predicted the
            correct initiation codon.  Where possible we choose an initiation
            codon (atg, gtg, ttg or (att)) which is preceded by an upstream
            ribosome binding site sequence (optimally 5-13bp before the
            initiation codon).  If this cannot be identified we choose the most
            upstream initiation codon.
            IMPORTANT: This sequence MAY NOT be the entire insert of the
            sequenced clone.  It may be shorter because we only sequence
            overlapping sections once, or longer, because we arrange for a
            small overlap between neighbouring submissions. Cosmid 2C4.
FEATURES             Location/Qualifiers
     source          1. .24225
                     /organism="Streptomyces coelicolor A3(2)"
                     /strain="A3(2)"
                     /db_xref="taxon:100226"
                     /clone="cosmid 2C4"
     misc_feature    complement(2. .1477)
                     /note="previously sequenced DNA fragment. EMBL:AF113603
                     Streptomyces coelicolor putative acyl-CoA carboxylase
                     complex A subunit (accA1) gene, complete cds."
     misc_feature    2. .119
                     /note="nominal overlap with Streptomyces coelicolor cosmid
                     SCAH10"
     misc_feature    complement(374. .1471)
                     /note="Pfam match to entry PF00289 CPSase_L_chain,
                     Carbamoyl-phosphate synthase (CPSase), score 585.30,
                     E-value 3.8e-172"
     misc_feature    complement(602. .625)
                     /note="PS00867 Carbamoyl-phosphate synthase subdomain
                     signature 2"
     RBS             complement(1485. .1489)
     RBS             1774. .1778
     gene            1786. .3438
                     /gene="SC2C4.02"
                     /note="SCO6272"
     CDS             1786. .3438
                     /gene="SC2C4.02"
                     /note="SC2C4.02, possible secreted FAD-binding protein,
                     len: 550 aa; similar to TR:Q9ZAR8 (EMBL:U80222)
                     Streptomyces peucetius ceasius DnrW, 485 aa; fasta scores:
                     opt: 1323 z-score: 1315.4 E(): 0; 45.3% identity in 470 aa
                     overlap. Contains Pfam match to entry PF01565
                     FAD_binding_4, FAD binding domain. Also contains possible
                     N-terminal region signal peptide sequence"
                     /codon_start=1
                     /transl_table=11
                     /product="putative secreted FAD-binding protein"
                     /protein_id="CAC22143.1"
                     /db_xref="GI:12274800"
                     /db_xref="SPTREMBL:Q9EX55"
                     /translation="MTEVSRRKLMKGAAVSGGALALPALGAPPATAAPAAGPEDLPGP
                     AAAAAGRRPAPAPVHVTRSDPRYQGLISGSNQRWVGNPDYIRVVSSAGQVVRAVQEAV
                     DKGLKIAVRSGGHCDEDFVANRDVRVVIDMAGMDSVTYDRERRAFAVGPGARLGTVYR
                     TLYKRWGVVLPGGTCPTVGAGGHITGGGYGALSRSRGLTVDHLYAVEVVHVDARGRAR
                     KVVATREEDDPNRELWWAHTGAGGGNFGVITRYWLRSPDAVGDDPSTLLPAAPSEVLL
                     SDVSWSWDDLDEASFTRLLRNFTEWHARNSAPDSPGRFLFSQLKTMHKAAGYFRMSTQ
                     VDAAAPDADRLLDDYLAAISEGTGVTYHVGDRYRAPWLYAVTEWSGFVEASVPRWKSK
                     SAYVREVMPEEQLRAVYRQLTRDDYPGPYGMIAIVGFGGKINEVAPGDTATAQRDSIA
                     KMLYCSLWSDPADDALHQRWIREAYEDVYASTGGVPRPGGVNDGCYINYADADLADPA
                     LNRSGIPWHELYFKGNYPRLQRVKATWDPRNVFSHRLGIRKP"
misc_feature         1948. .2535
                     /gene="SC2C4.02"
                     /note="Pfam match to entry PF01565 FAD_binding_4, FAD
                     binding domain, score 6.50, E-value 1.1e-06"
gene                 complement(3514. .9972)
                     /gene="SC2C4.03c"
                     /note="SCO6273"
CDS                  complement(3514. .9972)
                     /gene="SC2C4.03c"
                     /note="SC2C4.03c, probable type I polyketide synthase,
                     len: 2152 aa; N-terminal domain almost identical to
                     TR:Q9L8J0 (EMBL:AF202898) Streptomyces coelicolor A3(2)
                     type I polyketide synthase (fragment), 720 aa; fasta
                     scores: opt: 4356 z-score: 4494.6 E(): 0; 93.8% identity
                     in 721 aa overlap and C-terminal domain similar to
                     SW:HETM_ANASP (EMBL:L22883) Anabaena sp. polyketide
                     synthase HetM, 506 aa; blastp scores: Score= 761 (267.9
                     bits), Expect= 2.2e-74, Sum P(2)= 2.2e-74; Identities=
                     184/491 (37%), Positives= 278/491 (56%). Also highly
                     similar, in this same cosmid, to the N-terminal region of
                     SC2C4.04c and to SC2C4.05c. Contains Pfam matches to
                     entries PF00109 ketoacyl-synt, Beta-ketoacyl synthase,
                     PF00698 Acyl_transf, Acyl transferase domain and PF00550
                     pp-binding, Phosphopantetheine attachment site and matches
                     to Prosite entries PS00606 Beta-ketoacyl synthases active
                     site, PS00340 Growth factor and cytokines receptors family
                     signature 2 and PS00012 Phosphopantetheine attachment
                     site"
                     /codon_start=1
                     /transl_table=11
                     /product="putative type I polyketide synthase"
                     /protein_id="CAC22144.1"
                     /db_xref="GI:12274801"
                     /db_xref="SPTREMBL:Q9EX54"
                     /translation="MSNEQKMVEYLKWVTTDLQKARQRLAELEAGNDEPVAIVGMACR
                     FPGGVASPDDLWRLVMDEADGITDWPTDRGWDIDAIYDPERGKPGKSYTRKGGFLDGA
                     AQFDPGFFGISPREALAMDPQQRVLLETAWETFEQAGIDPATLRGSQVGVFAGVVEQS
                     YLGLEGPEELEGYLMTSKLSSVASGRIAYTFGFEGPAVSVDTACSSSLVALHLAVQSV
                     RKGESTMALAGGSTVSGNPSGFVDFSRQNGLAEDGRIKSFAAAADGTAWSEGVGLLLV
                     EKLSDARRNGHRVLAVVRGSAVNQDGASNGLTAPNGPSQERVIHQALAGARLNASDVD
                     VVEAHGTGTRLGDPIEAQALLASYGRQRPADRPLYLGSLKSNIGHTVAAAGAGGVIKM
                     IMAMRHGILPKTLNVDEPTPMVDWEAGAVELLTEARPWPETGAPRRAGVSAFGVSGTN
                     AHVILEEPPAEEPEEPADTAVTALPVLPWALSARTAPALRDQARRLLSHVEERPAEDP
                     LDVAYSLAAERSALEHRAVVVGSDREELLTGLRALAEGRTTPTGAVQDTRQTGKTAFL
                     FTGQGAQRTGMGMDLYRTFPAYAHAFDTITAQLDPHLDQPLHHTITTGHHLHHTGNTQ
                     PALFATEVALYRLLETWGITPDYLAGHSIGELTAAHISGILTLQDACTLVTARARLMQ
                     NLPTHGTMIALQATEQEILPHLTGHEHHLTIAAINSPTSLVISGNQTAAQHIATQLQA
                     QGRKTKTLTVSHAFHSPHMDGMLHDFHHTATQLTYHEPTIPIVSTLTGNLATHNDLRT
                     PTYWTDQLRNTVRYTQAIHTLHTAGVTTYTEIGPDATLTPLTQQTLDTTTAIPLLRHN
                     QPETHTLLTAISRLHNRGIPINWKNLFTTTPAKHTPLPTYAFQHERFWVETSGVAADA
                     SDLGLESAGHPMLGAAVTVAGADQVLFTSRLSLRTHPWLADHVVLDQTLVPATALVEM
                     AIRAGDELGSTVLNDLSVLGPLVLPGKGGVHVQVSVGSPDASGLRELSIYSRPDNADV
                     PWTLNARGRLSFRGPDAPFSLAQWPPAGAEEVPLDEAYDKLAVQGYVYGPVYRGLEHV
```

```
                     WRLGDEVYADVRLPDGSRTGIGGFGLHPALLETALAASCLAGYGSSPEGAVWISTDWK
                     DVRLHATGADRLRVRLAPDADGALNVQLADRAGRPVASVSSVACRAIAVGEVSNALAR
                     NRDSLFHVDWAPFPVAPSAGTERWVSLGADEPGARRYPDVASLARAVAAGESVDVALV
                     GTETGDTGDRVADVHDVTREALALVQDWLAEERLAGTRLVVRTRGGVTTAGDEDVTDL
                     GAAAVWGLLRSAQSENPGRIVLVDVDDEEASARILPAVVASGEPQSAVRDGQVFVPRL
                     GRVPEAPAGAATGAWDRRGTVLITGGTGSLGGLFARHLVVEHGVRNLLLTSRRGRDAE
                     GAAELEAELTALGARVEIAACDAADRAALAAVLADIPAEYPLTGVLHTAGVLDDGLIG
                     SQTPQRLSAVLRPKVDATWNLHELTRDMDLSAFVLFSSIAAVVGGPGQSTYSAANHFL
                     DALAEHRRSRGLAATSLAWGLWSQTSGMTGHLDEADLKRIARSGFRPVESEQGPALLD
                     IALTVGRAGLVATPMDLTALREQVQVPVLLSKLARTPVRRTAQNAEVSSESLAERLAG
                     LGDAEQREAVLGIVLGQAAVVLGHSDATGIDRQRSFTALGFDSLTSVELRNQLGTSTG
                     LRLPPTLVFDHPTPVALAEYLHGEVRAELSADGSASRPGVDFAAEVRLADDVRPADEV
                     ITTAADPKEILLTGASGFLGAFLLRDLMRTTTARVHCLVRGADEAAAMERLKANAEWY
                     RVWDDIDPERVSVVLGDLAEPRLGLDEESFDALARTVDVVYHNGARVHWLLPYETLKA
                     SNVTGTEEVLRLAARHRTVPVHYVSTVGVFDGVREPGVPLRVTDPTGPAESLPSGYLR
                     SKWVAEQVIEVARDRGLPVSVYRVDVISGDRVNGACQTRDFVWLTLKGLIQARSVPKG
                     TEGRFHLLPVDYVSAAITGISRQPGTVGRTFHLFNQSSLALSQCVELLRSLGYELDEV
                     DWDTWTQVVTSGDNALLPLLDAFEMMTSDTDGFYPPIDTAETVAALEGTGIGIPVLTR
                     ELFEKYVAFFVEEGHFPAAG"
     misc_feature    complement(4732. .4935)
                     /gene="SC2C4.03c"
                     /note="Pfam match to entry PF00550 pp-binding,
                     Phosphopantetheine attachment site, score 68.50, E-value
                     2.5e-17"
     misc_feature    complement(4810. .4857)
                     /gene="SC2C4.03c"
                     /note="PS00012 Phosphopantetheine attachment site"
     misc_feature    complement(5248. .5268)
                     /gene="SC2C4.03c"
                     /note="PS00340 Growth factor and cytokines receptors
                     family signature 2"
     misc_feature    complement(7342. .8286)
                     /gene="SC2C4.03c"
                     /note="Pfam match to entry PF00698 Acyl_transf, Acyl
                     transferase domain, score 248.50, E-value 9.5e-71"
     misc_feature    complement(8596. .9867)
                     /gene="SC2C4.03c"

  Query Match             48.4%;  Score 447.4;  DB 1;  Length 24225;
  Best Local Similarity   71.9%;  Pred. No. 5.9e-39;
  Matches  659;  Conservative    0;  Mismatches  216;  Indels   42;  Gaps    4;

Qy          1 ACCGCATTCATCTGCTCCGGACAGGGCACCCAACGCCCCGGCATGGCCCACGGCCTCTAC 60
              |||||| ||| ||| | |||| |||||| |||||||| |||||||| |    | |||||||
Db      13772 ACCGCGTTCCTCTTCACCGGCCAGGGCGCCCAACGCACCGGCATGGGCATGGACCTCTAC 13713

Qy         61 CACACCCACCCCGTCTTCGCCGCCGCACTCAACGACATCTGCACCCACCTCGACCCCCAC 120
              | ||||  ||||| || ||||  |||  || ||  ||||  | |||| ||||||||||||
Db      13712 CGCACCTTCCCCGCCTACGCCCACGCCTTCGACACCATCACCGCCCAACTCGACCCCCAC 13653

Qy        121 CTCGACCACCCCCTCCTCCCCCTCCTCACCCAAAACGACAACGACAACGAGGACGCGGCC 180
              |||||||| ||||||| || |  | |||||     | || ||
Db      13652 CTCGACCAACCCCTCCACCACACCATCACCACCGGCCACCAC------------------ 13611

Qy        181 GCACTGCTCCAGCAGACCCGCTACGCCCAGCCCGCCCTCTTCGCCTTCCAGGTCGCCCTC 240
                  ||||| || ||| || || |||| ||||||||||||||  | | |||||||||
Db      13610 ------CTCCACCACACCGGCAACACCCAACCCGCCCTCTTCGCCACCGAAGTCGCCCTC 13557

Qy        241 CACCGCCTCCTCACCGACGGCTACCACATCACCCCCCACTACTACGCCGGACACTCCCTC 300
              |||| |||||   |||   ||    |||||||||| |||||  |||||||||||||| ||
Db      13556 TACCGACTCCT---CGAAACCTGGGGCATCACCCCCGACTACCTCGCCGGACACTCCATC 13500

Qy        301 GGCGAAATCACCGCCGCCCACCTCGCCGGCATCCTCACCCTCACCGACGCCACCACCCTC 360
              |||||| |||||||||||||| || ||||||||||||||||   ||||||  |||||||
Db      13499 GGCGAACTCACCGCCGCCCACATCTCCGGCATCCTCACCCTCCAAGACGCCTGCACCCTC 13440

Qy        361 ATCACCCAACGCGCCACCCTCATGCAAACCATGC---CCCCCGGCACCATGACCACCCTC 417
              |||||  |||||||  |||||||||||| | |   ||| |||||||||||  | |||||
Db      13439 GTCACCGCACGCGCCCGCCTCATGCAAAACCTCCCCACCCACGGCACCATGATCGCCCTC 13380

Qy        418 CACACCACCCCCCACCACATCACCCACCACCTCACCGCCCACGAAAACGACCTCGCCATC 477
              ||  |||||   ||  | ||   || ||||||||||  |||||| || ||||| |||||
Db      13379 CAAGCCACCGAACAAGAAATACTCCCCCACCTCACCGGACACGAACACCACCTCACCATC 13320

Qy        478 GCCGCCATCAACACCCCCACCTCCCTCGTCATCAGCGGCACCCCCCCACACCGTCCAACAC 537
              |||||||||||| |||||||  |||||||||  ||||| ||    | ||| |||||||
Db      13319 GCCGCCATCAACAGCCCCACCAGCCTCGTCATCTCCGGCAACCAAACCGCCGCCCAACAC 13260

Qy        538 ATCACCACCCTCTGCCAACAACAAGGCATCAAAACCAAAACCCTCCCCACCAACCACGCC 597
              ||  ||||||    |||    |||||   |||||||||||||||| ||  |  |||||||
Db      13259 ATAGCCACCCAACTCCAGGCCCAAGGACGCAAAACCAAAACCCTCACCGTCTCCCACGCC 13200

Qy        598 TTCCACTCCCCCCACACCAACCCCATCCTCAACCAACTCCACCAGCACACCCAAACCCTC 657
              ||||||||||||||||   ||  ||| ||| || |  |||||||   | ||      |||
Db      13199 TTCCACTCCCCCCACATGGACGGCATGCTCCACGACTTCCACCACACCGCCACCCAACTC 13140

Qy        658 ACCTACCACCCACCCCACACCCCCCTCATCACCGCCAACA-----------CCCCACCC 705
              |||||||||  ||||  || |||| || || || || ||            || | | |
Db      13139 ACCTACCACGAACCCACCATCCCCATCGTCTCCACCCTCACCGGAAACCTCGCCACCCAC 13080

Qy        706 GACCAACTCCTCACCCCCCACTACTGGACCCAACAAGCCCGCAACACCGTCGACTACGCC 765
              || | |||| |||||||  |||||||||| | |||  |||||||||||||  |||| ||
Db      13079 AACGACCTCCGCACCCCCACCTACTGGACCGACCAACTCCGCAACACCGTCCGCTACACC 13020

Qy        766 ACCACCACCCAAACCCTCCACCAACACGGCGTCACCACCTACATCGAACTCGGACCCGAC 825
                 ||| ||| |||||||||      |||||||||||||||| |||| |||| ||||||
Db      13019 CAAGCCATCCACACCCTCCACACCGCCGGCGTCACCACCTACACCGAAATCGGCCCCGAC 12960

Qy        826 AACACCCTCACCACCCTCACCCACCACAACCTCCCCAACCCCCCCACCACCACCCTCACC 885
              |||||||||| |||||||||| || | ||||  || | || || ||| | |||||
Db      12959 GCCACCCTCACCCCCCTCACCCAGCAAACCCTCGACACCACCACCGCCATCCCCCTCCTA 12900

Qy        886 CTCACCCACCCCCACCA 902
              | |  | |||  | | |
Db      12899 CGCCACAACCAACCCGA 12883


RESULT 8
AF007101
LOCUS       AF007101               32870 bp    DNA     linear   BCT 13-JAN-1998
DEFINITION  Streptomyces hygroscopicus putative pteridine-dependent
            dioxygenase, PKS modules 1,2,3 and 4, and putative regulatory
            protein genes, complete cds and putative hydroxylase gene, partial
            cds.
ACCESSION   AF007101
VERSION     AF007101.1  GI:2624946
KEYWORDS    .
SOURCE      Streptomyces hygroscopicus.
  ORGANISM  Streptomyces hygroscopicus
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1  (bases 1 to 32870)
  AUTHORS   Ruan,X., Stassi,D., Lax,S.A. and Katz,L.
  TITLE     A second type-I PKS gene cluster isolated from Streptomyces
            hygroscopicus ATCC 29253, a rapamycin-producing strain
  JOURNAL   Gene 203 (1), 1-9 (1997)
  MEDLINE   98085969
   PUBMED   9426000
REFERENCE   2  (bases 1 to 32870)
  AUTHORS   Ruan,X. and Lax,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-JUN-1997) 47PAP9A, Abbott Laboratories, 100 Abbott
            Park Road, Abbott Park, IL 60064, USA
FEATURES             Location/Qualifiers
     source          1. .32870
                     /organism="Streptomyces hygroscopicus"
                     /strain="ATCC 29253"
                     /db_xref="taxon:1912"
     CDS             complement(108. .1130)
                     /note="orf5; similar to RapK"
                     /codon_start=1
                     /transl_table=11
                     /product="putative pteridine-dependent dioxygenase"
                     /protein_id="AAC38060.1"
                     /db_xref="GI:2624947"
                     /translation="MNPSSLVLNGLTSYFENGRARVVPPVGRNILGVVNYASVCEYPT
```

LDHGYPELEINMVAPTAEPFAEVWVTDAESEHGERDGITYAHDGEYFFCAGRVPPTGR
YTEATRAAYVTMPELLEBFGYSSVFRMWNFIGDINRDNAEGMEVYRDFCRGRAEAFEQ
CRLEFDQFPAATGIGSRGGGIAFYLLACRSGGHVHIENPRQVPAYHYPKRYGPRAPRF
ARATYLPSRAADGVGGQVFVSGTASVLGHETAHEGDLVKQCRLALENIELVISGGNLA
AHGISAGHGLTALRNIKVYVRRSEDVPAVREICREAFSPDADIVYLTVDVCRSDLLVE
IEGVVM"

CDS 1454. .9625
/note="orf1; loading domain"
/codon_start=1
/transl_table=11
/product="PKS module 1"
/protein_id="AAC38061.1"
/db_xref="GI:2624948"
/translation="MLRRELIRPLPELLKAHAARRGSQTAFADSRRGVSYAQLEARTR
RLAGHLAQLRLLPGDRAAIYLGNCVETIESYLAIARASAIGVPINPHSSTAELAYLLD
DSAARVIVTDHTRLPQVRRLQAERPHLTVVVTGDEESGAVPFDALAETEPQQPARDDL
GLDDIAWMLYTSGTTGRPKGVLSTQRSCLWSVAASSAGVLGLSETDRVLWPVPLFHSL
AHIYCVLSVTAVGATARITEAFDAEELLHTLRAEEFTFLAGVPTMYHYLLDAARDGDL
SLPNLRVCLSAGAISTATLRAEFQETFGVPLLDCYGSTETCGLITANWPNGTQAEGSC
GLPVPGLTVRIVDPETHESVETGEEGEIWVGGPSLMVGYHNQPEATEAALPRGWYRTG
DLGRRDDLGYHTITGRLSELIIRSGENIHPTEVEQVLLRVPGVTDAAVVGSAHPTLGE
VPVAFLVPAADGFDPDEVFAACREHLAYFKVPEELYEIGSIPRTGSGKIKRHALAQSP
ARLRAVSSGSFESLFRLEWTAPVPEAGPVDGDQATSAHRAGGSVTAGLEAPGSAVVDC
PRGSLDEVLRAVRTGLTDRSGRVAVLTHGADHDPAQAAVWGLVESMRADAPHDAPVLV
DADGGEATAALSAVLASGQPRALVRGGEVLVPRLARAGSGQRVSRPDPDGTVLVFGAD
QEQAAELAGHLVAAYGIRHVRLLARTAVPDLVARLTELGASVTQTVRAPGDRASLAEA
FQTAGPPVTAVLQVAGPETLPTTLTGEVELGPLQDALALDELTREAHPALFVLSAPID
NGPAHAVFQAVANRRHAAGLPALALSWGPREELTSRERLALFDAATTVDDASVLAIRH
ELAASAVDTGAAGRLARLPAAERGRALLDLVLSEVGVITGGRPVHDRQSFRDIGMTST
SAVELRGRLAAATGLALPATLAFDYPNPDALAAYLGGRLFGDTSAAGPDTEPTDPAAP
SGEPIAIVGMACRLPGGVTSPEELWELLSSGGDGISGFPTDRGWDLEHLYHPDPDHPG
TSYVREGGFLHDAGDFDAALFGISPREALAMDPQQRLLLETTWEVLERAGIDPTSLRG
SRTGVFAGVMYHDYISRLRTVPPELEGYLSSGGSGSVVSGRVSYTFGFEGPAVSVDTA
CSSSLVALHLAVQALRTGECTLAVAGGVTVMASPSSYIEFSRQRGLSPDGRCKSFAAD
ADGAAWSEGAATLLVERLSDAQRLGHPVLGVVRGTAVNQDGASNGLTAPNGPSQQRVI
RDALTNAQLTATDIDLIEAHGTGTKLGDPIEAQAVLAMYGQDRDLPVWLGSLKSNIGH
TQAAAGVAGVMKAVLSMRHGVMPKSLHIDEPTPEVDWSAGTVELLAEARDWPGADERP
RRAGVSSFGVSGTNAHVILEQAPPAPEPTESAEPVEPSEAAGPSGPVGLELGPSGPVP
WVLSAQNETALKDLARQLVTETEGTVHQIAHSLITQRSTTLPHRAVLVGSTREEFATA
LEQITRHTPDGGKTVLVFPGQGSQWVGMGQRLWETSPLYRDTIHEIDTALSEYIDWSV
TDILTSQPGAPTWDRVDVIQPVLFTTMTALATLWQHHGLTPDAIIGHSQGEIAAAYTA
GALTLQDAARITTLRSKALLPLTGHGTMISVLAPPEKIQEIADPDHWNTRIWIAAHNA
PNTTTITGDTQAITHLTQQLSHHRIMRWQLPGVDFAGHSGHIDQLHNQLQKILAPITP
QKPTIPFYSTTDNTWITDTTLDADYWYRNLRQPVLFQPAINHLTNNGYTTYIETSPHP
TLTPSIQETNPNTTTIHTLHNNQNDTHALLTALGHAHTHGHPITWHTLIPPTKTTPLP
TYPFQHTRYWLNDKPTVDTQDATQLGLIPTPHPLVPATTTLAETGATILTGHINPAHH
PWTTDHAVNGTPLLPGTALVDMALHAGDHTDHTTLDELIIHTPITLTHPTTIQLTISP
PDNTNTHTLTVHTRNSPDGPWTAHATGTLTQHQHTPDPPQTTWPPTDAEPIPLDGTYQ
RLATTGLEYGPTFQGLDALWRRDSRLFADVRLPNTDHTDSADGYGIHPALLDAALHAL
VAVSDGDEIRLPFAWTGVTLHATGATKLRVELEWDASGSASLRAFDPSGQPVVTVESL
ASRVISPEQLRSASARSGPLYRPEWIAWSGAGTPGGAVEFERYEVPTVGDDVLADTHR
RSTETLTRIQQHLAADTTTPLVVEATHDDLAGAAVWGLIRTAQTEHPGRIILIDTDNH
PTSRKALPTLIASGEPQARIRDGAITLPRLTVVAPTEPEAGVGEGTVLITGGTGSLGSL
IARHLVTEHGVRDLVLVSRQGPDAPGATELIAELQQLGAQVRVRSCDTTDRAALAALV
DEVGPPAGVVHTAGALADATVDNLDADALATTIAAKAHAAWWLHEHTQDHGLTLFMVF
SSLASVLGSAGQANYAAANSFLDALATHRRTLGLPGTSLAWGSWERTSGMSAHLTRAD
HDRLARAGLRPLTDEQGLEFFDGALAVDLGQVVTATFDLPAIGRSGAVPVLLRGLVRA
PRRAAVGADSGGAEPWTERIATLAPDARERSVLTSVIEQVAQALGHPDPGAIEPDQAF
KSLGFDSLTAVEFRNRLSAMTGLQLPATLVFDYPAAGAVTRYLLDQVSPAPAAAGPDD
DPDERELRRVLERVPVARFREAGVLEALLKLAGPDPARTPDAAEPDEDGLIDAMDTEA
LIRHVMDGTGA"

CDS 9637. .14925
/note="orf2"
/codon_start=1
/transl_table=11
/product="PKS module 2"
/protein_id="AAC38062.1"
/db_xref="GI:2624949"
/translation="MTNEQLVEALRVSAKENARLRRENASLQEGAAEPVAIVGMACRL
PGGVDSPESLWELMASGGDAVSPFPTDRGWDLDGLYDPDPDHLGTAYVREGGFLDRAG
AFDPDPFGISPREALAMDPQQRLLLETTWEVLERAGIDPTTLRGSRTGVFTGIMYQDY
VTRLRDFPAELEGYLGNGSAASVASGRVSYTFGFEGPAVSLDTACSSSLVALHLAVRP
ALRRLLARVAGGVTVMSVPHPYVEFSRQRGLAFDARCKSFDASADGTVLSEGVGLLLV
ERLSDALRLGHPVLAVVRGTAVNQDGASNGLTAPTGPSQQRVIRDALTDARLTVADVD
LLEAHGTGTRLGDPIEAQALIATYGQDRDQPVWLGSLKSNIGHTQAAAGVAGVIKTVM
AMRHGVMPRTLHVSEPTPNVDWSAGEVRLLTWARDWPSTDQRPRRAAVSSFGISGTNA
HVVLEQAPAAPEPDVPVTPPAGPVPWVLSAHTGPALSALAERLAAHVEEDEAPLGDVA
HTLNTTRATLPHRAVVVGETREEFVAALRAVRPTRPRRGPLAVMFTGQGSQRPGMGRQ
LYEHFPVFAQALDEVFALATPGLREVMFDPDQAETLQRTDHAQIALFAPETALYRLWE
SWGLRPDMVCGHSVGEITAAHVSGTLTLPDAVHLVTTRGTLMQNLPPGGAMLAVATDP
HTLQPHLDNHHDTISIAAINGPHATVLSGDRTTLHHIATQLNTKTNWLNVSHAFHSPL
MQPILQPFTTTLNTLTHHPPHTPLISMLTATPTHPDTTHWTQHITAPVRYTDTLHHLH
HHGITTYLEIGPDTTLTALARTTLPTTTHLIPTTRRNHNEVRSTNEALGRVFSVGHSV
DWRALTPTGRRTSLPTYPFQRRDFWLHDAAGGGAEVEGAGLGTTDHPLLGAVTTVADT
GELILSGRLSTSTHPWLTDHTVNGTVIVPGTALIDLALHAAEHTDHTTVDELVIHTPL
ALTHPTPSQSTVGAETDTGHRPIALHSRDTTDGTWTRHTTGTLSNQTHPAAELSTWSPT
DAHQIDLTTAYQQLADTGLHYGPAFQGLHTLYQRDNTLFAEIELPEAAGLPTGHILHP
ALLDAALHPAAVSAARDEESGIQLPFSWAGVTVHSTGATHLRVRLDTDDTGAIRRLAT
DTTGQPVITIDSLTTRPLHPDQLHAAQPETNHLYTLGWSPTPITPTTTGHFERYTVPT
AGDDILADTHRITAQTLTRIQQHLAADTTTPLVVEATHDDLAGARVWGLIRTAQTEHP
GRIILIDTDNHPTSRDALASVVASGKPQARIRDGAITLPRLTRLSVDTRASTERRPDL
SMGPVLVTGGTGGLGSLIARHLVTEHGVLDLVLVSRQGPDADGATELTNQLQQHGARI
RTISCDLTDRTALTTLIDEIGPLTGVVHTAGALADTTIDHLNPDTLTTTLTPKANPAW
WLHELTQDHDLALFVVFSSVAGVLGNAGQANYAAANSFLDALITHRRRQGLPGTSLAW
GWWQREGGMTAHLTQADHDRMTRAGIHGLTDAEGTTLFDTALDRGLGAVAPVKLHLPT
LNRADTVPAVLRGLVRAVRPAARQATVPAESWADRIAALMPEERDRAVLDLVRQQSAR
VLGHSDADRIDPGQAFKDTGFDSLTAVEFRNRVTTATGLRLPATLIFDHPTPNALADH
LLAQLGPLGDGGSSSPEVVRDLERFERQLFDTRPDDEVGTAIAVRLRSILAHLDAPAT
APAARTASAQVLDTLEDASQSELLSFIDKEFGRASN"

CDS 14964. .21266
/note="orf3"
/codon_start=1
/transl_table=11
/product="PKS module 3"
/protein_id="AAC38063.1"
/db_xref="GI:2624950"
/translation="MADSKTLTDYLKWVTADLYRTRERLAEVESASAEPIAIIGMACR
FPGGVTSPEALWELVSSGTDAISGFPDDRGWDMEGFFDSDPAQEGTSYMREGGFLTNA
TGFDPAFFGISPREAVAMDPQQRLLLETTWEVLERAGIDPTSLRGSRTGVFTGLMANS
YTTRVRRIPGELESYAGNGEAASVASGRVSYTFGFEGPAMSVDTACSSSLTALHLAVR
ALRSGECPLAVAGGVTVFATPRIFSDLSRQRGLARDGRCKAFSTSADGTGLSEGVGLL
LVERLSDAQRLGHSVLGVVRGTAVNQDGASNGLTAPNGPAQQRVIKAALADARLTAAD
IDLIEAHGTGTKLGDPIEAQAIIATYGQDREQPLWMGSLKSNIGHSQAAAGVAGIIKI
VMAMRHGVMPKTLHIDEPTPEVDWSAGTVELLAEARDWPGVDERPRRAGVSSFGVSGT
NAHVILEQAPPAPEPTESAEPVEPSEAAGPSGPVGLELGPSGPVPWVLSAQNETALKD
LARQLATVDGSVRDIAHSLISTRAALPHRAVVVGSTREEFATALEQITGHTPDGGKTV
LVFPGQGSQWVGMGQRLWETSPLYRDTIHEIDTALSEYIDWSVTDILTSQPGAPTWDR
VDVIQPVLFTTMTALATLWQHHGLTPDAIIGHSQGEIAAAYTAGALTLQDAARITTLR
SKALLPLTGHGTMISVLAPPQTIQDIADPDHWNTRIWIAAHNAPNTTTITGDTQAITH
LTQQLSHHRIMRWQLPGVDFAGHSGHIDQLHNQLQKILAPITPQKPTIPFYSTTDNTW
ITDTTLDADYWYRNLRQPVLFQPAINHLTNNGYTTYIETSPHPTLTPSIQETNPNTTT
IHTLHSNQNDTHALLTALGHAHTHGHPITWHTLIPPTKTTPLPTYPFQHTHFWLNEKT
ADADVESAGLSRTDHPLLGATTTLAHSGETVLTGRLSLTHHGWIADHAVNGTPLLPGT
ALVDMALHAGDHTDHTTLDELIIHTPITLTHPTTIQLTTTPTHDDNTLNLTIYTRTAP
EQPWIQHAAGTLTRANTAPSSPETIWPPTNAEPIPLDGTYQRLATTGLEYGPAFQGLH
ALWRRDSDNRLLADVRLPDTIDNTNSTDCTDGYGIHPALLDAALHALVAVSDGDEIRL
PFAWTGVTLHATGATKLRVELDIDDDGTLSLSATDTSGQPVVTVDSLTTRPIDPAQLR
ARPRQTDDLYALDWLPWSGSVAPAPEFEQYEVPKVGDDVLADTHRITAQTLTRVQQHL
AADTTTPLVVEATHDDLAGAAVWGLIRTAQTEHPGRIILIDTDNHPTSRDALASVVAS
GEPQARIRDGAITLPRLTTTAVTDALTPPADADAWHLDTSERGTLENLTLTGSDTATR
NLLPGEVRVAVRAAGLNFRDVLIALGMYPGEATMGSEAAGIVLDIGPDITDLTPGDRV
TGLFPNALANTAVADRRMITPIPASWSNTDAAAFPIVYLTAYYGLHDLAALKPGERVL
IHAGTGGVGTAAIQLAQHLGAEVFATAHPTKWPHLRAMGLDDHHIATSRTPDFETTFH
TTTQGQGIHVILNSLTDDLTDASLRLLTPNGRFIEMGKTDVRDTEEVARTHPGVTYRA
FDIADDAGPDRIGAMLVELTGLFERGALRGLPVRCWDVRESVDAFRFMAQARHIGKVV
LTIPHTPDGTILITGGTGNLGSLIAHHLVTTHGARDLILTSRQGPNAPGATELTNQLQ
QHGARIRTISCDLTDRTALTTLIDEIGPLTGVVHTAGALADTTIDHLNPDTLTTTLTP
KANPAWWLHELTQDHDLALFVVFSSVAGVLGNAGQANYAAANSFLDALITHRRRQGLP
GTSLAWGWWQREGGMTAHLTQADHDRMTRAGIHGLTDAEGTTLFDTALDRGLGAVAPV
KLHLPTIARSGSVLPVLRALVRAARPTAGQAAVQDGTWADRVAAVAPEERRAVVLELV
RAQAATVLGHSGAEHVSPTQAFKDSGFDSLTAVEFRNRVTTATGLRLPATLIFDHPTP
NTLADHLLAQLAPAYTGPSLADLDGIEVTLRTLAEDGERRPAVVARLKQWLSTLGAAQ
PAADTALDLDAASAREIFDFLDGKAG"

CDS 21363. .27176
/note="orf4"
/codon_start=1

```
                     /transl_table=11
                     /product="PKS module 4"
                     /protein_id="AAC38064.1"
                     /db_xref="GI:2624951"
                     /translation="MANEDEFLHYLKKAAADLRDARQQIQELEARDREPIAIVSMACR
                     LPGGVDSPESLWDLVAAGRDGVSAFPIDRGWDLEGIYDPDPDHPGTTYARDGGFLDDA
                     AGFDPAFFGISPREALAMDPQQRLLLETTWEVLERAGIDPAALRGSRTGVFTGVIHND
                     YADWSADDTDQLEGYLGNGSAASVASGRVSYTFGFEGPAVSLDTACSSSLVALHLAIQ
                     ALRSGECTLAITGGVTVLASVAPFIEFSRQRGLADDGRCKAFAADANGFGMAEGIGLL
                     LVERLSDAQRLGHPVLAVVRGTAVNQDGASNGLTAPNGPSQQRVIRDALTNAQLTATD
                     IDLVEAHGTGTKLGDPIEAQAIIATYGQHRDQPVRLGSLKSNIGHAQAAAGVAGIIKT
                     VMAMRHGIMPRTLHIDEPTPEVDWTTGTVELLTESRDWPTTDQHPRRAAVSSFGVSGT
                     NAHVILEQAPTTPEPHASEAPPAGPVPWILSAHTEPALRELARRLVTRHEESVTDVAY
                     SLLHGRAALPCRAVMIGSTREDFAAELEALTCGSGRASKVVFVFPGQGSQWIGMGRGL
                     WESSAVFRDSVLRTDRALAEFVDWSAAAVLAGEPGTPDLDRVDVVQPVLFTVMVALAN
                     VWRSYGVEPAAVVGHSQGEIAAAHVAGGLSLRDAARVVALRSRALTALAGEGGMVVVQ
                     RSVAYVENLLRQWEGRLSVAVVNGPEAVVVSGQVAALEELLATEDRARRVAVDYASHS
```

Query Match 37.3%; Score 344.6; DB 1; Length 32870;
Best Local Similarity 64.9%; Pred. No. 3.5e-28;
Matches 594; Conservative 0; Mismatches 279; Indels 42; Gaps 4;

```
Qy          4 GCATTCATCTGCTCCGGACAGGGCACCCAACGCCCCGGCATGGCCCACGGCCTCTACCAC 63
              ||  | || | | ||||||||||| ||||||||||||||||||  |     | ||| |
Db      11284 GCGGTGATGTTCACCGGACAGGGCTCCCAACGCCCCGGCATGGGACGACAGTTGTACGAG 11343

Qy         64 ACCCACCCCGTCTTCGCCGCCGCACTCAACGACATCTGCACCCACCTCGACCCCCACCTC 123
                |  |||||||||||||   |||||  ||||  ||| | | | |  |  ||||   ||
Db      11344 CACTTCCCCGTCTTCGCCCAGGCACTGGACGAGGTCTTCGCACTCGCCACCCCCGGACTA 11403

Qy        124 GACCACCCCCTCCTCCCCCTCCTCACCCAAAACGACAACGACAACGAGGACGCGGCCGCA 183
               | |      |  ||  || |                                 |||  ||||| |
Db      11404 CGCGAGGTGATGTTCGACCCC--------------------------GACCAGGCCGAA 11436

Qy        184 CTGCTCCAGCAGACCCGCTACGCCCAGCCCGCCCTCTTCGCCTTCCAGGTCGCCCTCCAC 243
               |||||  |  |||  | ||||||||  |||||| |||||||||  |   |||||| ||
Db      11437 ACACTCCAACGCACCGACCACGCCCAGATCGCCCTGTTCGCCTTCGAAACCGCCCTCTAC 11496

Qy        244 CGCCTCCTCACCGACGGCTACCACATCACCCCCCACTACTACGCCGGACACTCCCTCGGC 303
               ||    |||   ||   ||    | |    ||| ||     |  |||||||||  ||||
Db      11497 CG---ACTCTGGGAATCCTGGGGCCTGCGACCCGACATGGTCTGCGGACACTCGGTCGGA 11553

Qy        304 GAAATCACCGCCGCCCACCTCGCCGGCATCCTCACCCTCACCGACGCCACCACCCTCATC 363
               |||||||||| |||||  || |||||| |||||||||| |||||||  |  |||| ||
Db      11554 GAAATCACCGCAGCCCACGTCTCCGGCACCCTCACCCTCCCCGACGCCGTCCACCTCGTC 11613

Qy        364 ACCCAACGCGCCACCCTCATGCAAACCATGCCCCC---CGGCACCATGACCACCCTCCAC 420
               |||  ||||| |||||||||||||| | |||||||   |||| |||||  | || ||  |
Db      11614 ACCACACGCGGCACCCTCATGCAAAACCTGCCCCCCGGCGGCGCCATGCTCGCCGTCGCC 11673

Qy        421 ACCACCCCCCACCACATCACCCACCACCTCACCGCCCACGAAAACGACCTCGCCATCGCC 480
               |||  |||||||  | ||   | |||||||  |  |||| |  ||  | || ||||||||
Db      11674 ACCGACCCCCACACCCTCCAACCCCACCTCGACAACCACCACGACACCATCTCCATCGCC 11733

Qy        481 GCCATCAACACCCCCACCTCCCTCGTCATCAGCGGCACCCCCCACACCGTCCAACACATC 540
               |||||||||  ||||  | ||  |||| ||  ||||  || |  ||||  |||| ||||||
Db      11734 GCCATCAACGGCCCCCACGCCACCGTCCTCTCCGGCGACCGCACCACCCTCCACCACATC 11793

Qy        541 ACCACCCTCTGCCAACAACAAGGCATCAAAACCAAAACCCTCCCCACCAACCACGCCTTC 600
               |||||         ||||    || |||||||||    |||  |  || ||||||||||
Db      11794 GCCACC---------CAACTCAACACCAAAACCAACTGGCTCAACGTCAGCCACGCCTTC 11844

Qy        601 CACTCCCCCCACACCAACCCCATCCTCAACCAACTCCACCAGCACACCCAAACCCTCACC 660
               ||||||||||  ||   | |||||||||  | |   ||  |     |  | | ||||||||||
Db      11845 CACTCCCCCCTCATGCAACCCATCCTCCAACCCTTCACCACCACCCTCAACACCCTCACC 11904

Qy        661 TACCACCCACCCCACACCCCCCTCATCACCGCCAACACCCCACCCGACCAACTCCTCACC 720
               |||||||  || ||||| ||||||||||  |      ||||  |   |    |  || |  |
Db      11905 CACCACCCCCCACACACACCCCTCATCAGCATGCTCACCGCCACACCCACCCACCCCGAC 11964

Qy        721 CCCCACTACTGGACCCAACAAGCCCGCAACACCGTCGACTACGCCACCACCACCCAAACC 780
               ||  | ||||||||||| ||    |  |     |||||  |||| ||  ||||  |||   |
Db      11965 ACCACCCACTGGACCCAGCACATCACCGCACCCGTCCGCTACACCGACACCCTCCACCAC 12024

Qy        781 CTCCACCAACACGGCGTCACCACCTACATCGAACTCGGACCCGACAACACCCTCACCACC 840
               |||||||| |||||| |||||||||||| ||||| |||| ||||||| |||||||||| ||
Db      12025 CTCCACCACCACGGCATCACCACCTACCTCGAAATCGGCCCCGACACCACCCTCACCGCC 12084

Qy        841 CTCACCCACCACAACCTCCCCAACCCCCCCACCACCACCCTCACCCTCACCCACCCCCAC 900
               ||| ||| |  || |||||||| | || ||  |  || || ||||  |  || |  ||||
Db      12085 CTCGCCCGCACCACCCTCCCCACCACCACCCACCTCATCCCCACCACCCGCCGCAACCAC 12144

Qy        901 CACCACCCCCAAACC 915
               || |   ||  | |
Db      12145 AACGAAGTCCGCAGC 12159
```

```
RESULT 9
SCU88833
LOCUS       SCU88833                927 bp    DNA     linear   BCT 01-DEC-1998
DEFINITION  Streptomyces coelicolor polyketide synthase homolog gene, partial
            cds.
ACCESSION   U88833
VERSION     U88833.1  GI:3935135
KEYWORDS    .
SOURCE      Streptomyces coelicolor A3(2).
  ORGANISM  Streptomyces coelicolor A3(2)
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1  (bases 1 to 927)
  AUTHORS   Kuczek,K., Pawlik,K., Kotowska,M. and Mordarski,M.
  TITLE     Streptomyces coelicolor DNA homologous with acyltransferase domains
            of type I polyketide synthase gene complex
  JOURNAL   FEMS Microbiol. Lett. 157 (1), 195-200 (1997)
  MEDLINE   98079450
   PUBMED   9418255
REFERENCE   2  (bases 1 to 927)
  AUTHORS   Kuczek,K., Pawlik,K., Kotowska,M. and Mordarski,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-FEB-1997) Microbiology, Institute of Immunology and
            Experimental Therapy, Polish Academy of Sciences, Weigla 12,
            Wroclaw 53-114, Poland
REFERENCE   3  (bases 1 to 927)
  AUTHORS   Kuczek,K., Pawlik,K., Kotowska,M. and Mordarski,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-DEC-1998) Microbiology, Institute of Immunology and
            Experimental Therapy, Polish Academy of Sciences, Weigla 12,
            Wroclaw 53-114, Poland
  REMARK    Sequence update by submitter
COMMENT     On Dec 1, 1998 this sequence version replaced gi:2522317.
FEATURES             Location/Qualifiers
     source          1. .927
                     /organism="Streptomyces coelicolor A3(2)"
                     /strain="A3(2) M145"
                     /db_xref="taxon:100226"
     CDS             <1. .>927
                     /note="similar to ketosynthase and acyltransferase domains
                     of type I polyketide synthases"
                     /codon_start=1
                     /transl_table=11
                     /product="polyketide synthase homolog"
                     /protein_id="AAC80578.1"
                     /db_xref="GI:3935136"
                     /translation="DCTNAHVILEEPPAEEPEEPADTLHGAPGTAVGAVGPDPGGRDR
                     TGARLLSFVESRPEAAPADIGLSLAVSRSQFDHRARSSADREELLAGLRTLASGATTQ
                     TAVRDTRQTGKTAFLFTGQGAQRTGMGMDLYRTFHAYAHAFDTITAQLDPHLDQPLHH
                     TITTGHHLHHTGNTQPALFATEVALYRLLETWGITPDYLAGHSIGELTAAHISGILTL
                     QDACTPVTARARLMQNLPTHGTMIALQATEQEILPHLTGHEHHLTIAAINSPTSLVIS
                     GNQTAAQHIATQLQAQGRKTKTLTVSHAFHSPH"
BASE COUNT      189 a    418 c    213 g    107 t
ORIGIN
```

Query Match 33.8%; Score 312.2; DB 1; Length 927;
Best Local Similarity 73.5%; Pred. No. 1.9e-24;
Matches 452; Conservative 0; Mismatches 133; Indels 30; Gaps 3;

```
Qy          1 ACCGCATTCATCTGCTCCGGACAGGGCACCCAACGCCCCGGCATGGCCCACGGCCTCTAC 60
              ||||| ||| ||| | |||| |||||| |||||||| ||||||||| |   | |||||||
Db        340 ACCGCGTTCCTCTTCACCGGCCAGGGCGCCCAACGCACCGGCATGGGCATGGACCTCTAC 399

Qy         61 CACACCCACCCCGTCTTCGCCGCCGCACTCAACGACATCTGCACCCACCTCGACCCCCAC 120
              | ||||  || ||  || ||||  |||  || ||  ||||  | |||| ||||||||||||
Db        400 CGCACCTTCCACGCCTACGCCCACGCCTTCGACACCATCACCGCCCAACTCGACCCCCAC 459

Qy        121 CTCGACCACCCCCTCCTCCCCCTCCTCACCCAAAACGACAACGACAACGAGGACGCGGCC 180
              |||||||| ||||||| || |  | |||||      | || ||
Db        460 CTCGACCAACCCCTCCACCACACCATCACCACCGGCCACCAC---------------- 501

Qy        181 GCACTGCTCCAGCAGACCCGCTACGCCCAGCCCGCCCTCTTCGCCTTCCAGGTCGCCCTC 240
                  ||||| || ||| || || |||| ||||||||||||||  | | ||||||||||
Db        502 ------CTCCACCACACCGGCAACACCCAACCCGCCCTCTTCGCCACCGAAGTCGCCCTC 555

Qy        241 CACCGCCTCCTCACCGACGGCTACCACATCACCCCCCACTACTACGCCGGACACTCCCTC 300
              |||| |||||   |||   ||    |||||||||| |||||  |||||||||||| ||
Db        556 TACCGACTCCT---CGAAACCTGGGGCATCACCCCCGACTACCTCGCCGGACACTCCATC 612

Qy        301 GGCGAAATCACCGCCGCCCACCTCGCCGGCATCCTCACCCTCACCGACGCCACCACCCTC 360
              |||||| |||||||||||||| || ||||||||||||||||||   ||||||  |||||
Db        613 GGCGAACTCACCGCCGCCCACATCTCCGGCATCCTCACCCTCCAAGACGCCTGCACCCCA 672

Qy        361 ATCACCCAACGCGCCACCCTCATGCAAACCATGC---CCCCCGGCACCATGACCACCCTC 417
              |||||  |||||||  |||||||||| | | |   ||| ||||||||||| | |||||
Db        673 GTCACCGCACGCGCCCGCCTCATGCAAAACCTCCCCACCCACGGCACCATGATCGCCCTC 732

Qy        418 CACACCACCCCCCACCACATCACCCACCACCTCACCGCCCACGAAAACGACCTCGCCATC 477
              ||  |||||   ||  | ||   || ||||||||||  |||||| || ||||| |||||
Db        733 CAAGCCACCGAACAAGAAATACTCCCCCACCTCACCGGACACGAACACCACCTCACCATC 792

Qy        478 GCCGCCATCAACACCCCCACCTCCCTCGTCATCAGCGGCACCCCCCACACCGTCCAACAC 537
              |||||||||||||  |||||||  |||||||||  |||||  ||    | |||  |||||||
Db        793 GCCGCCATCAACAGCCCCACCAGCCTCGTCATCTCCGGCAACCAAACCGCCGCCCAACAC 852

Qy        538 ATCACCACCCTCTGCCAACAACAAGGCATCAAAACCAAAACCCTCCCCACCAACCACGCC 597
              ||  ||||||    |||   |||||  ||||||||||||||||  ||  |  |||||||
Db        853 ATAGCCACCCAACTCCAGGCCCAAGGACGCAAAACCAAAACCCTCACCGTCTCCCACGCC 912

Qy        598 TTCCACTCCCCCCAC 612
              |||||||||||||||
Db        913 TTCCACTCCCCCCAC 927
```

RESULT 10
AF016585

```
LOCUS       AF016585               41097 bp    DNA     linear   BCT 07-DEC-1997
DEFINITION  Streptomyces caelestis cytochrome P-450 hydroxylase homolog (nidi)
            gene, partial cds; polyketide synthase modules 1 through 7 (nidA)
            genes, complete cds; and N-methyltransferase homolog gene, partial
            cds.
ACCESSION   AF016585
VERSION     AF016585.1  GI:2558836
KEYWORDS    .
SOURCE      Streptomyces caelestis.
  ORGANISM  Streptomyces caelestis
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1  (bases 1 to 41097)
  AUTHORS   Kakavas,S.J., Katz,L. and Stassi,D.
  TITLE     Identification and characterization of the niddamycin polyketide
            synthase genes from Streptomyces caelestis
  JOURNAL   J. Bacteriol. 179 (23), 7515-7522 (1997)
  MEDLINE   98053867
   PUBMED   9393718
REFERENCE   2  (bases 1 to 41097)
  AUTHORS   Kakavas,S. and Stassi,D.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-JUL-1997) 47p, Abbott Laboratories, 100 Abbott Park
            Road, Abbott Park, IL 60064, USA
FEATURES             Location/Qualifiers
     source          1. .41097
                     /organism="Streptomyces caelestis"
                     /strain="NRRL-2821"
                     /db_xref="taxon:36816"
     gene            complement(<1. .396)
                     /gene="nidi"
     CDS             complement(<1. .396)
                     /gene="nidi"
                     /note="ORF6; similar to tyli"
                     /codon_start=1
                     /transl_table=11
                     /product="cytochrome P-450 hydroxylase homolog"
                     /protein_id="AAC46023.1"
                     /db_xref="GI:2558837"
                     /translation="MVDSVTGPMELSKDANAKELLDWFSHNRTHHPVFWDEGRQAWQV
                     FRYDDYLTVSNHPEFFSSDFTEVAPTPPELEMILGPGTIGALDPPAHGPMRKLVSQAF
                     TPRRMAGQEQRIRVIAEELLDRVRGQKTIA"
     gene            627. .13649
                     /gene="nidA1"
     CDS             627. .13649
                     /gene="nidA1"
                     /note="orf1"
                     /codon_start=1
                     /transl_table=11
                     /product="polyketide synthase modules 1 and 2"
                     /protein_id="AAC46024.1"
                     /db_xref="GI:2558838"
                     /translation="MAGHGDATAQKAQDAEKSEDGSDAIAVIGMSCRFPGAPGTAEFW
                     QLLSSGADAVVTAADGRRRGTIDAPADFDAAFFGMSPREAAATDPQQRLVLELGWEAL
                     EDAGIVPESLRGEAASVFVGAMNDDYATLLHRAGAPTDTYTATGLQHSMIANRLSYFL
                     GLRGPSLVVDTGQSSSLVAVALAVESLRGGTSGIALAGGVNLVLAEEGSAAMERVGAL
                     SPDGRCHTFDARANGYVRGEGGAIVVLKPLADALADGDRVYCVVRGVATGNDGGGPGL
                     TVPDRAGQEAVLRAACDQAGVRPADVRPVELHGTGTPAGDPVEAEALGAVYGTGRPAN
                     EPLLVGSVKTNIGHLEGAAGIAGFVKAALCLHERALPASLNFETPNPAIPLERLRLKV
                     QTAHAALQPGTGGGPLLAGVSAFGMGGTNCHVVLEETPGGRQPAETGQADACLFSASP
                     MLLLSARSEQALRAQAARLREHLEDSGADPLDIAYSLATTRTRFEHRAAVPCGDPDRL
                     SSALAALAAGQTPRGVRIGSTDADGRLALLFTGQGAQHPGMGQELYTTDPHFAAALDE
                     VCEELQRCGTQNLREVMFTPDQPDLLDRTEYTQPALFALQTALYRTLTARGTQAHLVL
                     GHSVGEITAAHIAGVLDLPDAARLITARAHVMGQLPHGGAMLSVQAAEHDLDQLAHTH
                     GVEIAAVNGPTHCVLSGPRTALEETAQHLREQNVRHTWLKVSHAFHSALMDPMLGAFR
                     DTLNTLNYQPPTIPLISNLTGQIADPNHLCTPDYWIDHARHTVRFADAVQTAHHQGTT
                     TYLEIGPHPTLTTLLHHTLDNPTTIPTLHRERPEPETLTQAIAAVGVRTDGIDWAVLC
                     GASRPRRVELPTYAFQRRTHWAPGLTPNHAPADRPAAEPQRAMAVGPVSREALVRLVG
                     ETTASVLGLDGPDEVALDRPFTSQGLDSMTAVELAGLLGTAAGVALDPTLVYELPTPR
                     AVADHLAKTLLGESAADADQEVNGRTGEAEAKAGDPIAVIGIGCRFPGGVATPDDLWE
                     LVASGTDAISTFPTDRGWDLDGLYDPDPSTPGKSYVRHGGFLHDAAQFDAEFFGISPR
                     EATAMDPQQRLLLETSWEALERAGVVPESLRGGRTGVFVGTTAPEYGPRLHEGTDGYE
                     GFLLTGTTASVASGRIAYALGTRGPALTVDTACSSSLVALHLAVQSLRRGECDLALAG
                     GTTVMSGPGMFVEFSRQRGLAPDGRCKAFSADADGTAWAEGVGMLLVERLSDAERLGH
                     RVLAVVRGTAVNQDGASNGLTAPSGPAQQQVIRDALSDAGLSADDIDAVEAHGTGTAL
                     GDPIEAGALLATYGHPKRQTPVWLGSLKSNIGHTQAAAGIAGIIKMVQALRHDTLPRT
                     LHADHPSSKVDWDAGPLQLLTDARPWPADPDRPRRAGISAFGVSGTNAHVVLEEPPAP
                     EEAPARTQAPPPVIAWPLSAHTPTALRAQAARLDSWLHHTQADPQDIAHALATTRTHF
                     KHRAVITGRTRTELHTKLHTLDAIQGTAHPHPRLTLLFTGQGAQHRGMGQELYATDPH
                     FAAALDEVCEELQRCGTQNLREVMFTPDQPDLLDRTEYTQPALFALQTALYRTLTARG
                     TQAHLVLGHSVGEITAAHIAGVLDLPDAARLITARAHLMGQLPHGGAMLSVQAAEHDL
                     DQLAHTHGVEIAAVNGPTHCVLSGPRTALEETAQQLHQQGIRHTWLKVSHAFHSALMD
                     PMLGAFRDTLNTLNYQPPTIPLISNLTGQIADPNHLCTPDYWIDHARHTVRFADAVQT
                     AHDQRTTTYLEIGAHPTLTTLLHHTLDNPTTIPTLHREHPEPETLTTALATLHTTGHT
                     TTLHTTSPQTHHLDLPTYPFQRDRYWMEPVRVAQVSGQPGADRLRYRVVWEAAPEDHT
                     VPSAERWLLLGGSDPADAGILVGVEEQLVGHGAAVQRLDIPAPAGRRLLPSCSELPVR
                     GPSHTGEFCGSRLRSPRVAEAVALLQALGDAGADAPLWIATRGAVAAQAGEAPSVGGA
                     QLWGLGQVAGLELADRWGGLVDLPADPAPAALRGLARVLVSNAPDNQVAIRASGVFVR
                     RVVPAPGRPVRNDWAPSGTVLITGGTGALGSQVARRLALAGAPHLLLAGRRGNSGAAE
                     LVDELTALGAEVTVAACDAADRDALVSLLATIPEHRPLTAVLHAAGVLDDGVLDGLTP
                     ERIDAVLRAKATAARHLDELTADLDLDAFVLFSSIVGVWGNGGQATYAAANAALDALA
                     HGRRARGQRATSIAWGPWAGSGMAAGDGAKSLARDGIDVLEPEEALDLLEQAVRSDET
                     TLLVADVDWETFLSSSASRRFQLPLFGRIPAARAVAVSGADGSPDDVPVWLSPQLSAA
                     ERQLRLLDLVRTEAASVLRHGSATAVDAETAFRAGGFDSLTLVELRNRLASATGLRLS
                     NTLLFDHPTPAAVAEYLGEQLLGTEPSAAASHGDAGARTVVASDEPVAVVGMACRYPG
                     GVATPDDLWELVASGTDAISTFPTDRGWDLDGLYDPDPSTPGKSYVRHGGFLHDAAQF
                     DAEFFGISPREATAMDPQQRLLLETSWEALEHAGIAPHTLRTTRTGVFTGASQQEYGT
                     QSREAADKYGGHLLTGTLASVMSGRVAYTLGLQGPALTVDTACSSSLVALHLAVQSLR
                     RGECDLALAGGSTVMSTPTVFVEFSRQRGLAPDGRCKPFADTADGTAWAEGVGVLLVE
```

RLSDAERLGHRVLAVVRGTAVNQDGASNGLTAPSGPAQQQVIRDALSDAGLSADDIDA
VEAHGTGTPLGDPIEAEALLATYGHPERQTPVWLGSLKSNIGHSQAAAGIAGIIKMIQ
AMHHGTLPRTLHVDAPTSQVDWEAGSLQLLAEARPWPADPDRPRRAGISAFGVSGTNA
HVVLEEPPAPEEAPARTQAPPPVIAWPLSAHTPTALRAQAARLDSWLHHTQADPQDIA
HALATTCTHFKHRAVITGRTRTELHTKLHTLDAIQGTAHPHPRLTLLFTGQGAQHPGM
GQELYTTDPHFAAALDEICEELQRCGTQNLREVMFTPDQPDLLDRTEYTQPALFALQT
ALYRTLTAHGTQAHLVLGHSVGEITAAHIAGVLDLPDAARLITARAHLMGQLPHDGAM
LSVQAAEHDLDQLAHTHGVEIAAVNGPTHCVLSGPRTALEETAQHLREQNVRHTWLKV
SHAFHSALMDPMLGAFRDTLNTLNYQPPTIPLISNLTGQIADPNHLCTPDYWIDHARH
TVRFADAVQTAHDQRTTTYLEIGPHPTLTTLLHHTLDNPTTIPTLHREHPEPETLTTA
LATLHTTGHTTTPHPSHIPAQRVSLPAYPFQRRAYWMPNSAAHIGRSDAEAATRLGLA
RQDHPFLTGATSIVGSGATLLTGRVSLAAHPWLADHTVAGAVLFPGTAFADLLLHVAE
ETGSGTVEELTLHVPLVLPAQGAVQLQVLVEAPDERGRRAATVAARPEEETGAEWTTH
AAGVLAGAEPAAPAVDWADWAAGAWPPPGAEAVDVDELYAEFAAAGYDYGPAFTGLTG
AWRRGEEVFADVRLPEAVAGSGEEERRFGVHPALFDAALHPWRVGGPWQESGPDADGG
QSTLLPFSWQGLTLYGTGAEQVRVRLAPSDGGAFSVRAADPAGHPVLALDALVLRRVP
SSTPGALPAAPMYHIDWQPVDSTSLVDTRRYAVVGADADETAEVLRAFLPDAHTTAHA
DLAALRAAPDAVAPTPGPALVVALIPGSPPPAPEATSAPDPVRATLAHGLSLVQALIE
DERLTTGTTLAVVTRGAVEATQTDVPDLAGAALWGLVRSAQSEYPGRFRLIDVDDSPE
SRAALATALASDEPQLALRAGVMLAPALTPAEAAGSPAPADAAPTAHGMDGTVLLTGC
TGALGRRVAVHLARRHGVRRMLLVSRSGPNAPESAALERELTELGVHTSFVSCDLADP
ADVRKAVAGVPSDHPLTGVVHTAGVLDDGALARLTPERIDTVLRAKADAVRHLHEATL
GHPLRAFVLFSAAAGLLGRPGQASYAAANAVLDAFARLRRAEGLPAVSLAWGLWDERE
GMAGGLDATALRRLRRDGILPMPADQALELLDQALTTYRDGPALLVPLLLDSTALRRT
AGENGPASVPPLLRSLVPSHPRRHRTTTATSVRDESGPGPSARLAALPAKERMTVLLD
TVSEQVAGVLGHASAAEIDPERPFHEIGFDSLATLELRNRLGRLVGLRLPSTLAFDHP
TPRKVAEWINAELPAANPALGTSAEDPVLTGIDQMARAIALMDAEDGRRADVRQRLVG
LLTALDRPASGADSTRAAATTVADRLDEATDDEIFAFLDEQLEEQPDEPLGAPLDERL
DEQP"

gene 13693. .19212
/gene="nidA2"

CDS 13693. .19212
/gene="nidA2"
/note="orf2"
/codon_start=1
/transl_table=11
/product="polyketide synthase module 3"
/protein_id="AAC46025.1"
/db_xref="GI:2558839"
/translation="MTAENDKIRSYLKRATAELHRTKERLAELESARHEPVAVVGMAC
RYPGGVATPDDLWELVASGTDAISTFPTDRGWDLDGLYDPDPSTPGKSYVRHGGFLHD
AAQFDAEFFGISPREATAMDPQQRLLLETSWEALEHAGIAPNTLRTTRTGVFTGVMYD
DYGSRCFPSLPEVEGYVVNGSAGSVASGRVAYTLGLQGPALTVDTACSSSLVALHLAV
QSLRRGECDLALAGGSTVMSTPTVFVEFSRQRGLAPDGRCKSFADTADGTAWAEGVGV
LLVERLSDAERLGHRVLAVVRGTAVNQDGASNGLTAPSGPAQQQVIRDALSDAGLSAD
DIDAVEAHGTGTPLGDPIEAGALLATYGHPERQTPVWLGSLKSNIGHTQAAAGIAGII
KMIQAMHHGTLPRTLHVDHPSSKVDWKAGSLQLLAQARPWPADPDRPRRAAVSAFGVS
GTNAHAIIEEPSTLIEAPVPDETPVPAHTPPVIAWPLSAHTPTALRAQAARLDSWLHH
TQADPQDIAHALATTRTHFKHRAVITGRTRTELHTKLHTLDAIQGTAHPHPRLTLLFT
GQGAQHPGMGQELYTTDPHFAAALDEVCEELQRCGTQNLREVMFTPDQPDLLDRTEYT
QPALFALQTALYRTLTARGTQAHLVLGHSVGEITAAHIAGVLDLPDAARLITARAHLM
GQLPHGGAMLSVQAAEHDLDQLAHTHGVEIAAVNGPTHCVLSGPRTALEETAQHLREQ
NVRHTWLKVSHAFHSALMDPMLGAFRDTLNTLNYQPPTIPLISNLTGQIADPNHLCTP
DYWIDHARHTVRFADAVQTAHHQGTTTYLEIGPHPTLTTLLHHTLDNPTTIPTLHREH
PEPETLTTALATLHTTGHTTTLHTTSPQSHHLDLPTYPFQRDRYWMAVPPRAAVGDLA
AAGVTAVGHPLLAAVTEVPETDTTVWTGMLSARALPWLADHLVWDRGVAPGTAVLEIV
LEIGERIGLPCIGELVLEAPLSWGADAPAVVRTVVGALSPEGTREVSLYSRPEDEAVW
TRHASGRLARAGSEEMGSGVFAELSGVWPPPGAQEVDISGQYELFAAAGVGYGGAFRG
LRAVWRREDEIFAEVRLPDDHATDAAGYGLHPALLDAALHPIALLDPLGDGGHGLLPF
VWTDVLRPGNGGHTLRVRVTRAQDPADGGQDDRPGGAVSLAAADRDGVPVFVARSLVL
RRVSATQLPAAPFVPLHRLDWVPIPGFTESTGAGGSDPSVLDLTELTADVHLTVREVL
EAVQAHLREPVGDGGRLVVLTRGAAGPEGGADAPAGAAAWGLVRSAQSEHPGRFVLID
VDGTEESSRALPAVMAAVGDVDAPQVAVRQGRPSVPRLVPLTADRTEAEPGGPGSAGL
LGPDGTVLITGGTGALAAVTARHLVARHGARHLLLVSRRGTDAPQAAELASELGALGA
DVEFRACDVTDRAAVQRLLAGIPAEHPLSCVVHTAGVLDDGVISAQTPERVAAVLGPK
ADGALHLDELTRDRERSVPLVLFSSVSGVLGSTGHAGYAAANAFLDALAARRRADGHP
ALSLGWGWWTGTGMAAGLDRADTARIRRSGIAPLDAETAMDLFDRALARPEPALLPVR
LDHAALREAARAGTLPEVLRDLAGVPSRAHPAPGSEGTATSTGTGSGLAERLARRTVA
ERLPLLLELVRTEVAAVLGHGDPGAVGAERSFKDAGFDSLTAVDLRNRLNARTGLRLP
ATLVFDHPTPLALAELLLAGLDVEPAADGPEDLAAAFDRLARGLAALDDRAARVGAAE
RLRSLLTALTSADPATGTGAQAGPEADGALTEVVTDRLRSASDDELFDLYDSEFR"

gene 19269. .30233
/gene="nidA3"

CDS 19269. .30233
/gene="nidA3"
/note="orf3"
/codon_start=1
/transl_table=11
/product="polyketide synthase modules 4 and 5"
/protein_id="AAC46026.1"
/db_xref="GI:2558840"
/translation="MSATNEEKLREYLRRAMGDLHSARERLAELESARHEPVAVVGMA
CRYPGGVATPDDLWELVASGTDAISTFPTDRGWDLDGLYDPDPSTPGKSYVRHGGFLH
DAAQFDAEFFGISPREATAMDPQQRLLLETSWEALEHAGIAPHTLRTTRTGVFTGVMH
HDYGSHQVGSAADASGQLGLGTAGSVASGRVAYTLGLQGPALTVDTACSSSLVALHLA
VQSLRRGECDLALAGGTTVMATPTVFVEFSRQRGLAPDGRCKPFADTADGTAWAEGVG
MLLVERLSDAERLGHRVLAVVRGTAVNQDGASNGLTAPSGPAQQQVIRDALADAHLTP
DDIDAVEAHGTGTPLGDPIEAGALLATYGHPKRQTPVWLGSLKSNIGHTQAAAGIAGI
IKMVQALRHDTLPRTLHADHPSSKVDWEAGPLQLLAQARPWPADPDRPRRAGISAFGV
SGTNAHAIIEEPPTLTEAPADTHTPPVIAWPLSAHTPTALRAQAARLRARLHRTSDAL
SLADSAGIGHGLAVGRAALPHRAVLLGDGAAPLDALAALASGEVSPDVVTGSAADVRR
VAFVFPGQGAQWAGMGAELLDSSPVFAAELARCEAALEPFVDWSLTDVLRGAPGAPGL
DRVDVVQPVTFAVVVALAAMWRWLGVEPAAVVGHSQGEIAAAHVAGVLSLEDAARVVA
LRSQLIARELAGRGSMASVALAAADVESRLAGAEAGGGVRDVEIAAVNGPETTVVCGA
PGAVDSLLGVLQGEGVRVRRIDVDYASHSRHVEGIRDELAAVLAGLRPRAGRVPFYST
VEAEPLDGTALDAGYWYRNLRQRVRFESALRAMLADGVDAFVECSPHPVLTVPVRQTL
EDAGAGAVAVGSLRRDDGGLRRFLTSAAEAQVAGVPVDWAALCPRAGWVDLPTYAFQR
ERYWVAPAEPGPAAGAGSAAATGPAAATAARSGSGQEDAASAYGDRLAYHVAWQGLLP
GADAWRPGPRLLIVPAGEREAEVADSVEQAIASFGGTVRRVTADPLSTGRDGLRALLA
PVVAGEEPLTGAVSLLGLCTDGHPEHPGVPAGVTGTLALVQALADLGVAVPLWTVTRG
TVLVVGDSTTVPARLLRSLLDDGAERVVLAGADAGTATDAGTAELAPDAAVVRIPCDV

Query Match 33.4%; Score 308.4; DB 1; Length 41097;
Best Local Similarity 62.8%; Pred. No. 2.1e-24;
Matches 578; Conservative 0; Mismatches 306; Indels 36; Gaps 5;

```
Qy      11 TCTGCTCCGGACAGGGCACCCAACGCCCCGGCATGGCCCACGGCCTCTACCACACCCACC 70
           ||| | |||| || ||  |||| | || |||||||| ||| |  |||||  |||  |||
Db    5218 TCTTCACCGGCCAAGGAGCCCAGCACCGCGGCATGGGCCAAGAGCTCTACGCCACAGACC 5277

Qy      71 CCGTCTTCGCCGCCGCACTCAACGACATCTGCACCCACCTCGACCCCCACCTCGACCACC 130
           |   |||||||||||| ||  |||| |||||    | ||  | | |  |  | |||
Db    5278 CGCACTTCGCCGCCGCCCTGGACGAAGTCTGCGAAGAACTACAGCGCTGCGGCACCCAGA 5337

Qy     131 CCCTCCTCCCCCTCCTCACCCAAAACGACAACGACAACGAGGACGCGGCCGCACTGCTCC 190
           ||| | |               | |  |    || |  |||  | | |  |||||
Db    5338 ACCTGCGCG--------------AGGTGATGTTCACCCCCGACCAGCCGGACCTGCTGG 5382

Qy     191 AGCAGACCCGCTACGCCCAGCCCGCCCTCTTCGCCTTCCAGGTCGCCCTCCACCGCCTCC 250
           | |  |||   ||| |||||||||| |||||||| | ||   |||||||  |||||  ||
Db    5383 ACCGCACCGAATACACCCAGCCCGCACTCTTCGCCCTGCAAACCGCCCTCTACCGCACCC 5442

Qy     251 TCACCGACGGCTACCACATCACCCCCCACTACTACGCCGGACACTCCCTCGGCGAAATCA 310
           | || | | |   |  ||      |||||  | |   || ||||||  ||||||||||||
Db    5443 TGACAGCCCG---CGGCACACAAGCCCACCTCGTCCTGGGCCACTCCGTCGGCGAAATCA 5499

Qy     311 CCGCCGCCCACCTCGCCGGCATCCTCACCCTCACCGACGCCACCACCCTCATCACCCAAC 370
           ||||  |||||| |||||||| |||||  |||  ||||||||  |   ||| ||||||  ||
Db    5500 CCGCAGCCCACATCGCCGGCGTCCTCGACCTGCCCGACGCCGCACGCCTGATCACCGCAC 5559

Qy     371 GCGCCACCCTCATGCAAACCATGCCCCCCGGCACCATGACCACCCTCCACACCACCCCCC 430
           ||||  |||||||  |    | ||||  |||| |  |    |  | |  |  |  || |
Db    5560 GCGCACACCTCATGGGACAACTCCCCCACGGCGGCGCCATGCTCTCCGTCCAAGCCGCAG 5619

Qy     431 ACCACATCACCCACCACCTCACCGCCCACGAAAACGACCTCGCCATCGCCGCCATCAACA 490
           | |||  |  | |||| ||   ||| |||    ||| | |||  |||||||| |||||
Db    5620 AACACGACCTCGACCAACT---CGCACACACCCACGGCGTCGAGATCGCCGCCGTCAACG 5676

Qy     491 CCCCCACCTCCCTCGTCATCAGCGGCACCCCCCACACCGTCCAACACATCACCACCCTCT 550
           ||||||  |  |||| |  |||  ||| |  | || |  || | | | ||   |
Db    5677 GACCCACCCACTGCGTCCTGTCCGGACCCCGCACCGCCCTGGAAGAAACCGCCCAGCAAC 5736

Qy     551 GCCAACAACAAGGCATCAAAACCAAAACCCTCCCCACCAACCACGCCTTCCACTCCCCCC 610
           || ||||| |||||      ||      ||          |  ||||||||||||||| |||
Db    5737 TACACCAACAGGGCATACGCCACACCTGGCTGAAGGTCTCCCACGCCTTCCACTCCGCCC 5796
```

```
Qy        611 ACACCAACCCCCATCCTCAACCAACTCCACCAGCACACCCAAACCCTCACCTACCACCCAC 670
              ||   ||||||| |||  |    |||  |  |   |  | | |||||||  |||||| || |
Db       5797 TCATGGACCCCATGCTCGGCGCCTTCCGCGACACCCTCAACACCCTCAACTACCAACCCC 5856

Qy        671 CCCACACCCCCCTCATCACCGCC----------AACACCCCACCCGACCAACTCCTCA 718
              ||  || ||||||||||| ||  |            | | ||   ||| |||| |||  ||
Db       5857 CCACCATCCCCCTCATCTCCAACCTCACCGGCCAGATCGCCGACCCCAACCACCTCTGCA 5916

Qy        719 CCCCCCACTACTGGACCCAACAAGCCCGCAACACCGTCGACTACGCCACCACCACCCAAA 778
              ||||| |||||||||| | | || || ||| |||||||  || ||||  | ||  ||| |
Db       5917 CCCCCGACTACTGGATCGACCACGCACGCCACACCGTCCGCTTCGCCGACGCCGTCCAGA 5976

Qy        779 CCCTCCACCAACACGGCGTCACCACCTACATCGAACTCGGACCCGACAACACCCTCACCA 838
              ||  |||| | ||  ||  |||||||||| ||||| |||   || ||  ||||||||||
Db       5977 CCGCCCACGACCAACGCACCACCACCTACCTCGAAATCGGGGCCCACCCCACCCTCACCA 6036

Qy        839 CCCTCACCCACCACAACCTCCCCAACCCCCCCACCACCACCCTCACCCTCACCCACCCCC 898
              |||||  |||||||| ||||  |||   |||||||||| || |||||||  || |   |
Db       6037 CCCTCCTCCACCACACCCTCGACAA---CCCCACCACCATCCCCACCCTCCACCGCGAAC 6093

Qy        899 ACCACCACCCCCAAACCCAC 918
              ||| | | ||  |||||| |
Db       6094 ACCCCGAACCAGAAACCCTC 6113
```

```
RESULT 11
SNA132222
LOCUS       SNA132222              20394 bp    DNA     linear   BCT 24-APR-1999
DEFINITION  Streptomyces natalensis pimS1 gene.
ACCESSION   AJ132222
VERSION     AJ132222.1  GI:4678702
KEYWORDS    pimaricin biosynthesis; pimS1 gene; polyketide synthase.
SOURCE      Streptomyces natalensis.
  ORGANISM  Streptomyces natalensis
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1  (bases 1 to 20394)
  AUTHORS   Aparicio,J.F., Colina,A.J., Ceballos,E. and Martin,J.F.
  TITLE     The biosynthetic gene cluster for the 26-membered ring polyene
            macrolide pimaricin. A new polyketide synthase organization encoded
            by two subclusters separated by functionalization genes
  JOURNAL   J. Biol. Chem. 274 (15), 10133-10139 (1999)
  MEDLINE   99214571
   PUBMED   10187796
REFERENCE   2  (bases 1 to 20394)
  AUTHORS   Aparicio,J.F.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-JAN-1999) Aparicio J.F., Microbiology, INBIOTEC, Av.
            Real, 1-Leon, 24006, SPAIN
FEATURES             Location/Qualifiers
     source          1. .20394
                     /organism="Streptomyces natalensis"
                     /db_xref="taxon:68242"
     gene            1. .20394
                     /gene="pimS1"
     CDS             1. .20394
                     /gene="pimS1"
                     /function="pimaricin biosynthesis"
                     /codon_start=1
                     /transl_table=11
                     /product="polyketide synthase"
                     /protein_id="CAB41041.1"
                     /db_xref="GI:4678703"
                     /translation="MSNEEKLREYLKRAIADLHETRQQLDETEAKQREPLAIVSMACR
                     FPGGVRSPEELWELLRDGVDAVSSFPRNRGWDLDALYHSDPAHQGTSYAREGGFLHDA
                     GEFDPGFFGISPREALAMDPQQRLLLETAWEAVERAGIDPESLAGSRTGVFVGTGHGG
                     YDAEGRRRADEVGGHLLTGNHISIASGRISYVLGLEGPALTVDTACSSSLVALHLAMH
                     ALRRDECAMALVGGATVMSTPQMFVEFSRQRGLAPDGRCKPFAAAADGTGWSEGVGLL
                     LVERLSDAVRNGYPVLAVLKGSAVNQDGASNGLTAPNGPSQQRVIRQALTGAGLAASD
                     IDAVEAHGTGTTLGDPVEAHALLATYGQQRAADRPCGLGSMKSNIGHTQAAAGIAGVM
                     KMVLAMRHGHLPRTLHLDEPTGHVDWSEGNARLLAEPEPWPSAGRPRRAAVSSFGISG
                     TNAHVILEQAPAHEAEPAPEPAARPGALPWILSARTEAGLRAQADRLGRHLRDRADLE
                     PAAVAHALADTRTLMEHRAVVVAGDREEFLRGLDALAAGRTANGLVSGVAVKAASAFL
                     FAGQGSQRPGMGRELHAAHPVFATAFDAVCAELDPHLDRPLRDIVFAEEGSAEAALLD
                     QTAYTQAALFALETALFRLVESWGVAPRFVAGHSIGELTAAHVSGVLTLHDAARLVAA
                     RGTLMQALPAGGAMVAVQATEDEIRERLAGHEDHVALAAANGPDSTVISGDEQAVTEI
                     AAHWEAQGRRTKRLRVSHAFHSPHMDDMLEDFRRVARGLTFHAPRIPVVSTVTGALAT
                     EDELRSPDYWVRQVRETVRFCAAVRTLEAEGVTTFVEIGTGGVLTPMVQDCLTTLEEP
                     VLVPLLRTGRPETVALTEGVATAFVHGVPVDRSAFPGAPGTSRADLPTYAFQRQWYWL
                     DPADHDEGEAAAAEAGEAGFWAAVEREDLQELSAVLAIDGSEADSLGSFLPTLSSWRR
                     QRRTQAAADRFSYRTHWAPRTASGGPTATGHWLVVLPEGGTDDPWTARLLDALNDQGL
                     HTDVRELPADHEPDAWGRHPVDGVLCLLALDERPTRSCPPYRRGLAATTNAAARPEGA
                     GIQAPLWCVTRGAVAVDRHEALKSPLQAQTWGLGRVAALESPQSWGGLIDLPDNLDGR
                     AVSALLSTLAGEEDQVAVRPAGVFARRLERITPGGDTGDRWSTHGTVLVTGGTGALGA
                     HLAHWLADAGAEHLVLTGRRGPQAPGAPELAAALTDRGVKVTLAACDAADRDALAAVL
                     ADIPPHLPLTGVVHAAGVLDDGVLDALTPERFETVLRPKARAAQNLHELTQDLDLDHF
                     VLFSSIVGVLGNAGQANYAAANAYLDALAEHRLAQGLPATSVSWGPGQAAAWHDSDAA
                     DRMSRDGLLPMAAAPRRRPAPALAQGMTQVTVADIDWSAYAPALTAVRPSPLIGDLPE
                     ARRALGPAEGPRRERSPLRDRIGALPPAEQEKAFLTMVREEAARVLGHPSPDTVDAQR
                     AFREQGFDSLMAVDLRNRLSAATGLRLPATLLFDHPTPLAAAACLRSEVLGAAGPATV
                     VQASTAALDEPVAIIGMACRFPGGVHSPEALWRLLAEGGDAITPMPADRGWDLDRLYH
                     PDPDHQGTSYARGGGFLDGAADFDADFFGISPREALAMDPQQRLLLETWEVLEQAGID
                     PESLRGSSTGVFAGTNTQDYGTALDAAQDEAGGHRLTGNAMSVVSGRVSYTFGFEGPA
                     LTVDTACSSSLVALHMAAQALRQGECSLAVAGGVTVMATPSSFVEFARQRGLAPDGRC
                     KPFAAAADGTGWSEGVGLLLVERLSDARRNGHQVLAVVRGSAVNQDGASNGLSAPSGP
                     SQQRVIRQALANARVAASEVDAVEAHGTGTTLGDPIEAQALLATYGQERPLLLGAVKS
                     NLGHTQAAAGVAGVMKMVLAMRHGMLPRTLHVDEPTGHVDWTAGAVELLTEHTDWPET
                     GHPRRAAVSAFGISGTNAHVVLELPAAEQPLVEQPSAAEPDAPATAPDRTPTASDGTA
                     PLLLSAKSESALRAQAARLHSHLERDPALRLTDAAYTLMTHRTAFAHRAAVRAADHEA
                     ALRALTALAAGEADPAVDTGTAHTGRDAVLFSGQGSQRIGMGRELSGRYPVFAEAFDT
                     VCAALDEHLDRPLRDVVRGEDEELLNRTVYAQAGLFAIEVALFRLVESWGVRPHYVAG
                     HSVGEIAAAHVAGVFSLADACALVAARGRLMQALPAGGAMAAIRATEDEVLPHLADSV
                     SIAAVNGPSSVVVSGAEHAVLSIAAHFEGAGRKTTRLRVSHAFHSPLMDPMLADFRAV
                     AEGLTYGEPELAVVSNVTGQLATPDQLRTPEYWVTHVRAAVRFADGIRALGAEGVTRF
                     LELGPDGVLSALARESAPDDAVCTPVLRKDRSEAATLLAALTHLHVHGTEIDWTAFLA
                     GRDAHAVDLPTYAFQHQRFWPTPDHTRTGDLGAVGLEATGHPLLSAAVELPDGEGLLF
                     TTRLSLQTHPWLAGHVVMGSVLLPGTAFAELALRAADEVGCDRVDELTLAAPLVLPEH
                     GGVQLQLRVGPADASGRRTLTARSRAEGDGDRPWVQHATGVLAEGESTPEPGYDFHTE
                     SWPPADAAPVELSGLYPDFAAHGFDYGPHFQGLRTAWRRGDEVFAEVALPAEAEGEAS
                     AYGLHPALLDAALHVVAFNGVDRGVVPPSWESVALHATGASAVRIRVVRHSGDTVSVD
                     VADTTGEPVASIGTLVLRAVSADQLAGGADPAVRDALFRVQWNPVRLPPAGAAVTVAT
                     LGSLAGAPFDGYPDLASLARSGRVAGAVLVPVEAGAGEVVADDVVGATHATAARALDL
                     ARSWLADDRFAASRLVFVTRGAVSGADLAGAAVWGLVRSALSEHPGRFGLVDLDDDAE
                     LALVPRVLASDEPQLLVRGGEVLAARLARAQSSHAVTWDPSGTVLVTGGTGGLGRVMA
                     RHLVVEHGVRNLLLVSRRGPAAEGAEELVTELRHSGAEVAVEACDVTDAAAVADLVAR
                     HRISAVVHTAGVLDDGVVESLTPERLSAVLRPKVDAAWNLHEATRDLDLDAFVVFSSV
                     AGTIGSPGQANYAAGNAFLDALAHHRRAAGLPAASLAWGPWSRDGGMTGTLTDVDSSA
                     SPGRHARTHPRTGRGLFDAALAAGDAHLLPVRFDWASLRAQGEVPPLLRGLIRTRARR
                     SAVGGSAAAAGLVGRLSGRGTVERREVLLDLVRAQIAVVLGHANPETIESTRVFQDLG
                     FDSLTAVELRNRLNNATGLRLSATAVFDYPTADALVDFLLDELFGAQEEAELPAPVPS
                     PAGAADDPVVIVGMSCRYPGGVGSPEDLWRLVSEGVDAVSDFPTDRGWDVESLYSPDP
                     EALGTSYTRSGGFLHEAAEFDPDFFGMSPREALATDAQQRLLLETTWEAIERTGIDPA
                     SLRGSRTGVFAGVMYTDYGDLLVGDQFEGYRSNGSAASIASGRVSYTFGFEGPAVTVD
                     TACSSSLVALHWAAQSLRSGECSLAVAGGVTVMSTPTTFVEFSRQRGLSADGRCKAPA
                     DAADGVGWGEGVGMLVLERLSDARRNGHRVLAVVRGSAVNQDGASNGLTAPNGPAQQR
                     VIRQALASAGLSAADVDAVEAHGTGTTLGDPIEAQALLATYGQERPEDRPLLLGSVKS
                     NIGHAQAASGVAGVIKMVLAMRHGVLPRTLHVDEPSSHVDWSAGAVELLTSEAEWPQG
                     EGPRRAGVSSFGVSGTNAHVILEQPGPDAADAAPDATVTDPGALAWVLSARNEAALRC
                     QAARLLSLVAGSDALCARDIGHSLVTGRSSFAHRAVVWGQDRDALVRALSALAVGEAD
                     AGLAEGASGAGRTAFLFSGQGSQRLGMGWELYARYPVFADAFDAVCAALDEHLERPLR
                     DVVWGEDAELLNQTAYAQAGLFAIEVALYRLAESWGMRPDFVAGHSIGEVAAAHVSGV
                     FSLPDACALVAARGRLMQQLPSGGAMMAIRATEDEVLPHLAEGVSLAAVNGPSSVVIS
                     GAEDAVLAIAAHFAGEGRKTTRLRVSHAFHSPLMEPMLEEFRAVVTRLSFGTPTIPVV
                     SNLTGRLAEPEQLAHADYWVRHVREAVRFADGIQALRAEGVTRFLELGPDGVLSAMAR
                     ESASDDAVLAPVLRRDRPEETALLGALAQLYVRGAHVDWTVPFAGSGARWADLPTYAF
                     QHERFWPSGGVARPGDVRSAGLGSAGHPLLGAAVELAGSGGLLFTGRLSVSSHPWLAD
                     HVVLGSVLVPGTALVELVLRAADEAGCDLLEELTLAAPLVLPASGAAVQVQVAVGEPD
                     EAGRRPVSVHARGEGPWTLHASGAVTSGAEVPPFDATVWPPKGAEPVDVADCYDVLA
                     DAGLTYGPAFHGLQAAWKLGGDVYAEAKLPESTDGDAYGLHPALFDAALHASALGGAE
                     AGGVPFSWAGVSLHATGASHLRVRIREAGGALSVAIADTSGAPVASVESLVIRPLSAG
                     QVQAADRDALFKADWVPVPLTDERVEPGTGPEGEPLRTYADLDSLEGAAVPGTVLVAP
                     PSGAAGTVESVHAATVWALEMVQAWLADDRFATSRLVFVTRGAAFGADLAAAAVRGLV
                     RSAQSENPGRFGLVDMDGDADTTVPAQALATDEPELLVRGGEVLAARLVRAQSSHTVT
                     WDPSGTVLITGGTGGLGRSVARHLVSEHGVRSLLLVSRRGPAAEGAGELVAELRGSGA
                     EVVIEACDVTDAVAVADLVARHRISAVVHTAGVLDDGVVESLTPERLAVVLRPKVDAA
                     WNLHEATRGLDLDAFVVFSSVAGTFGSAGQANYAAGNAFLDALAYHRRAVGLPAVSLA
```

```
                     WGPWSQDGGMTGTLSDADVQRIARQGMPPLTVEEGLALFDAALGSAEPMALPVRLDLA
                     ALRAQGEPQPLLRGLIRTRTRRSGAAAASGIAQRLAGLSTAERREALLDVVRAQIATV
                     LGHAGPETIAPDRAFQDLGLDSLTAIELRNLLGKATGLRLPATTVFDYPTVDALAAHL
                     LDELFGAETGTATETPLPVPGLPSLADDPVVIVGMSCRFPGGVASPEDLWRLVADGVD
                     AVSAFPTDRGWEIDDTYDPEREGAIATRSGGFLHDAAEFDPEFFGMSPREALTTDAQQ
                     RLLLETTWEALERAGMDPATLRGSRTGVFAGVMYHDYSTLLSGREFEGYQGSGSAGSV
                     ASGRVSYTFGFEGPAVTVDTACSSSLVALHLAAQSLRSGECSLALAGGVTVMSTPLTF
                     VEFSRQGGLSADGRCKAFADAADGVGWAEGAGILVLERLSDARRNGHRILATVRGSAV
                     NQDGASNGLTAPNGPAQQRVIRQALASAGLSAADVDAVEAHGTGTTLGDPIEAQALLA
                     TYGQERPEDRPLLLGSVKSNIGHAQAASGVAGVIKMVLAMRHGVLPRTLHVDEPSSHV
                     DWSAGAVELLTSEAEWPQGEGPRRAGVSSFGISGTNAHVILEQPEPVAAETESITPDT
                     APDAAEDEAADSGTPVPALLSGRSASALRAQAARLLSRLDGDPGPRITDVAYSLATGR
                     SAFPHRAVILAANRADLLHSLSALAEGHTEAPAVVAQDRARSGKLAFLFSGQGSQRLG
                     MGRELYGRYPAFAEALDAVCAALDAHLDRPLRDVIWGEDAELLNRTGYAQTGLFAIEV
                     ALFRLLESWGVRPDHLLGHSIGEIAAAHVAGVLSLPDACALVAARGRLMQQLPSGGAM
                     MAIRATEDEVLPHLAEGVSLAAVNGPSSVVVSGAEDEVLALAAHFEEEGRKTTRLRVS
                     HAFHSPLMEPMLADFRAVADGMTYAAPRIPVVSNVTGRPATAEELCCAEYWVGHVREA
                     VRFADGVGALREQGVTTFLELGPDGSLSALAAESAADDSVLAPVLRKNRPEAPALLTA
                     LARLHAQGTPVDWSAAFAGTGARWVDLPTYAFQHERFWPSGGAARAGDVRSAGLGSAG
                     HPLLGAAVELAGSGGRLLTGRLSLSSHPWLADHVVLGSVLVPGTALMELVLRAADEVD
                     CAAVDELTLAAPLVLPASGAAIQVQVWVGEPDEAGRRPVSVHAREGEGPWTLHADGAL
                     APAAETVPFDTAIWPPQGAEHLDAAGCYERFADAGFAYGPVFQGLRAAWKLGEDIYAE
                     VALPEGTDGNAYGLHPALFDAALHAALLGGEGTDEAAVPFSWNGVTLHATGASRVRVR
                     IRPTEGGTSIALVDTAGAPVASVRSLTARPITAGQLQTGDRDSLFQVDWTTLHLTDER
                     ANSLALLGKDTEGILDTLSLQPHADLDDLAATGVHDTVLAPLPTRTAGTVESVHAATT
                     GALALIRSWLADDRFAASRLVFVTRGAVSGTDLAGASVWGLVRSALLEHPGRFGLVDV
                     DVDQDAEVPLVPRALASDEPQVLVRGGEVLAARLVRAQSSDTVTWDPSGTVLITGGTG
                     GLGRSVARHLVSEHGVRSLLLVSRRGPAAEGVDALVAELAECGAQVTVEACDVTDAVA
                     VADLVARHRISAVVHTAGVLDDGVVESLTPERLSAVLRPKVDAAWNLHEATRGLDLDA
                     FVVFSSVAGTFGSAGQANYAAGNAFLDALAYHRRAVGLPAVSLAWGPWSQDGGMTGTL
                     SDADVQRIARQGMPPLTVEEGLALFDAALGSAEPMALPVRLDLAALRAQGEPQPLLRG
                     LIRTPGRRTAAAATEGDTAAAFAGRLTGLSAAEGREVVLGAVRSQIAGVLGHAEATEI
                     DQDRAFLDLGFDSLTAVELRNRLGAVTGIRLPATLLFDYPTPAELVAHLHARIAPEPT
                     VGPEALLGELERMEKSFGGLDLTEEMHEQIAGRLEVLRAKWDALRDTAAAAGHDGSPS
                     DEDFDFESASDDEVFDLLDNELGLS"
BASE COUNT     2611 a   7486 c   7397 g   2900 t
ORIGIN

  Query Match           32.7%;  Score 302.2;  DB 1;  Length 20394;
  Best Local Similarity   60.9%;  Pred. No. 1.1e-23;
  Matches  572;  Conservative    0;  Mismatches  343;  Indels   24;  Gaps    4;

Qy          1 ACCGCATTCATCTGCTCCGGACAGGGCACCCAACGCCCCGGCATGGCCCACGGCCTCTAC 60
              | ||| ||| ||| | |||| ||||| |||| || || ||||||| | ||  ||  ||
Db       1687 AGCGCGTTCCTCTTCGCCGGGCAGGGCTCCCAGCGACCGGGCATGGGGCGCGAACTGCAC 1746

Qy         61 CACACCCACCCCGTCTTCGCCGCCGCACTCAACGACATCTGCACCCACCTCGACCCCCAC 120
                | | |||||||| |||||| | ||  || |||   | ||| || | || ||||| |||
Db       1747 GCCGCGCACCCCGTGTTCGCCACGGCCTTCGACGCGGTGTGCGCCGAACTGGACCCACAC 1806

Qy        121 CTCGACCACCCCCTCCTCCCCCTCCTCACCCAAAACGACAACGACAACGAGGACGCGGCC 180
              || ||||  || || | |  | || ||        || | | ||   |   | || ||||
Db       1807 CTGGACCGGCCGCTGCGCGACATCGTC------TTCGCCGAGGAAGGCAGCGCCGAGGCC 1860

Qy        181 GCACTGCTCCAGCAGACCCGCTACGCCCAGCCCGCCCTCTTCGCCTTCCAGGTCGCCCTC 240
              || |||||| | |||||  |||| ||||| |||| |||||||||  |   |   ||||||
Db       1861 GCCCTGCTCGACCAGACCGCCTACACCCAGGCCGCGCTCTTCGCCCTGGAAACCGCCCTG 1920

Qy        241 CACCGCCTCCTCACCGACGGCTACCACATCACCCCCCACTACTACGCCGGACACTCCCTC 300
                |||  ||| |   |||   ||    | |  | ||||  | |  |||||||||||| ||
Db       1921 TTCCGGCTCGT---CGAATCCTGGGGCGTGGCACCCCGGTTCGTCGCCGGACACTCCATC 1977

Qy        301 GGCGAAATCACCGCCGCCCACCTCGCCGGCATCCTCACCCTCACCGACGCCACCACCCTC 360
              |||||  | |||||||||||| ||   ||| | || ||||||  ||||||| |   ||
Db       1978 GGCGAGCTGACCGCCGCCCACGTCAGTGGCGTGCTGACCCTCCACGACGCCGCACGGCTG 2037

Qy        361 ATCACCCAACGCGCCACCCTCATGCAAACCATGCCCCC---CGGCACCATGACCACCCTC 417
              || ||   |||| |||||||||||   |  ||||| |   |||| |||||    |  ||
Db       2038 GTCGCCGCGCGCGGCACCCTCATGCAGGCGCTGCCCGCAGGCGGCGCCATGGTGGCGGTC 2097

Qy        418 CACACCACCCCCCACCACATCACCCACCACCTCACCGCCCACGAAAACGACCTCGCCATC 477
              ||  |||||    || | |||  | | |  ||| ||| |||||| || || ||||| ||
Db       2098 CAGGCCACCGAGGACGAGATCCGCGAGCGTCTCGCCGGCCACGAGGACCACGTCGCCCTC 2157

Qy        478 GCCGCCATCAACACCCCCACCTCCCTCGTCATCAGCGGCACCCCCCACACCGTCCAACAC 537
              || |||  ||||   |||   |||  ||||||    |||  |  . ||  |||||   |
Db       2158 GCGGCCGCCAACGGGCCCGATTCCACCGTCATTTCGGGCGACGAACAGGCCGTCACCGAG 2217

Qy        538 ATCACCACCCTCTGCCAACAACAAGGCATCAAAACCAAAACCCTCCCCACCAACCACGCC 597
              ||| |  ||| |||   |   ||| |||  |   |||||     || |    || |||||||
Db       2218 ATCGCGGCCCACTGGGAGGCACAGGGCCGCCGCACCAAGCGGCTGCGGGTCAGCCACGCC 2277

Qy        598 TTCCACTCCCCCCACACCAACCCCATCCTCAACCAACTCCACCAGCACACCCAAACCCTC 657
              ||||||||||| ||||   ||  ||| ||  |  |  |||  |    | |||      ||
Db       2278 TTCCACTCCCCGCACATGGACGACATGCTGGAGGACTTCCGGCGCGTCGCCCGCGGTCTG 2337

Qy        658 ACCTACCACCCACCCCACACCCCC------------CTCATCACCGCCAACACCCCACCC 705
              |||| |||| | |||| || ||||               | | |  |||   | || |
Db       2338 ACCTTCCACGCCCCCCGCATCCCCGTGGTGTCCACGGTGACCGGCGCGCTCGCCACCGAA 2397

Qy        706 GACCAACTCCTCACCCCCCACTACTGGACCCAACAAGCCCGCAACACCGTCGACTACGCC 765
              ||| |||| | | | ||| ||||||||   |  || | |||| | ||||||  || |
Db       2398 GACGAACTGCGCTCGCCCGACTACTGGGTGCGGCAGGTCCGCGAAACCGTCCGCTTCTGT 2457

Qy        766 ACCACCACCCAAACCCTCCACCAACACGGCGTCACCACCTACATCGAACTCGGACCCGAC 825
              | ||   | ||||  |      | |||||||||||||| | | ||  |||| ||| |
Db       2458 GCTGCCGTGCGCACCCTTGAGGCCGAGGGCGTCACCACCTTCGTGGAGATCGGCACCGGC 2517

Qy        826 AACACCCTCACCACCCTCACCCACCACAACCTCCCCAACCCCCCCACCACCACCCTCACC 885
              |  |||||| || |  ||| ||  ||  ||| || |        ||   ||| |
Db       2518 GGCGTCCTCACCCCCATGGTCCAGGACTGTCTGACCACCCTCGAAGAGCCCGTTCTCGTC 2577

Qy        886 CTCACCCACCCCCACCACCACCCCCAAACCCACCTCCTC 924
              |     |  | |  |  || |||| ||||| |   |||
Db       2578 CCTCTGCTGCGCACCGGCCGCCCCGAAACCGTCGCTCTC 2616


RESULT 12
AX067996
LOCUS       AX067996               20394 bp    DNA     linear   PAT 19-JAN-2001
DEFINITION  Sequence 3 from Patent WO0077222.
ACCESSION   AX067996
VERSION     AX067996.1  GI:12329806
KEYWORDS    .
SOURCE      Streptomyces natalensis.
  ORGANISM  Streptomyces natalensis
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1  (bases 1 to 20394)
  AUTHORS   Martin,J.F., Aparicio,J.F. and Colina,A.J.
  TITLE     Genes encoding enzymes in the biosynthesis of pimaricin and the
            application thereof
  JOURNAL   Patent: WO 0077222-A 3 21-DEC-2000;
            DSM N.V. (NL)
FEATURES             Location/Qualifiers
     source          1. .20394
                     /organism="Streptomyces natalensis"
                     /db_xref="taxon:68242"
     CDS             1. .20394
                     /note="unnamed protein product"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAC24642.1"
                     /db_xref="GI:12329807"
                     /translation="MSNEEKLREYLKRAIADLHETRQQLDETEAKQREPLAIVSMACR
                     FPGGVRSPEELWELLRDGVDAVSSFPRNRGWDLDALYHSDPAHQGTSYAREGGFLHDA
                     GEFDPGFFGISPREALAMDPQQRLLLETAWEAVERAGIDPESLAGSRTGVFVGTGHGG
                     YDAEGRRRADEVGGHLLTGNHISIASGRISYVLGLEGPALTVDTACSSSLVALHLAMH
                     ALRRDECAMALVGGATVMSTPQMFVEFSRQRGLAPDGRCKPFAAAADGTGWSEGVGLL
                     LVERLSDAVRNGYPVLAVLKGSAVNQDGASNGLTAPNGPSQQRVIRQALTGAGLAASD
                     IDAVEAHGTGTTLGDPVEAHALLATYGQQRAADRPCGLGSMKSNIGHTQAAAGIAGVM
                     KMVLAMRHGHLPRTLHLDEPTGHVDWSEGNARLLAEPEPWPSAGRPRRAAVSSFGISG
                     TNAHVILEQAPAHEAEPAPEPAARPGALPWILSARTEAGLRAQADRLGRHLRDRADLE
                     PAAVAHALADTRTLMEHRAVVVAGDREEFLRGLDALAAGRTANGLVSGVAVKAASAFL
                     FAGQGSQRPGMGRELHAAHPVFATAFDAVCAELDPHLDRPLRDIVFAEEGSAEAALLD
                     QTAYTQAALFALETALFRLVESWGVAPRFVAGHSIGELTAAHVSGVLTLHDAARLVAA
```

RGTLMQALPAGGAMVAVQATEDEIRERLAGHEDHVALAAANGPDSTVISGDEQAVTEI
AAHWEAQGRRTKRLRVSHAFHSPHMDDMLEDFRRVARGLTFHAPRIPVVSTVTGALAT
EDELRSPDYWVRQVRETVRFCAAVRTLEAEGVTTFVEIGTGGVLTPMVQDCLTTLEEP
VLVPLLRTGRPETVALTEGVATAFVHGVPVDRSAFPGAPGTSRADLPTYAFQRQWYWL
DPADHDEGEAAAAEAGEAGFWAAVEREDLQELSAVLAIDGSEADSLGSFLPTLSSWRR
QRRTQAAADRFSYRTHWAPRTASGGPTATGHWLVVLPEGGTDDPWTARLLDALNDQGL
HTDVRELPADHEPDAWGRHPVDGVLCLLALDERPTRSCPPYRRGLAATTNAAARPEGA
GIQAPLWCVTRGAVAVDRHEALKSPLQAQTWGLGRVAALESPQSWGGLIDLPDNLDGR
AVSALLSTLAGEEDQVAVRPAGVFARRLERITPGGDTGDRWSTHGTVLVTGGTGALGA
HLAHWLADAGAEHLVLTGRRGPQAPGAPELAAALTDRGVKVTLAACDAADRDALAAVL
ADIPPHLPLTGVVHAAGVLDDGVLDALTPERFETVLRPKARAAQNLHELTQDLDLDHF
VLFSSIVGVLGNAGQANYAAANAYLDALAEHRLAQGLPATSVSWGPGQAAAWHDSDAA
DRMSRDGLLPMAAAPRRRPAPALAQGMTQVTVADIDWSAYAPALTAVRPSPLIGDLPE
ARRALGPAEGPRRERSPLRDRIGALPPAEQEKAFLTMVREEAARVLGHPSPDTVDAQR
AFREQGFDSLMAVDLRNRLSAATGLRLPATLLFDHPTPLAAAACLRSEVLGAAGPATV
VQASTAALDEPVAIIGMACRFPGGVHSPEALWRLLAEGGDAITPMPADRGWDLDRLYH
PDPDHQGTSYARGGGFLDGAADFDADFFGISPREALAMDPQQRLLLETWEVLEQAGID
PESLRGSSTGVFAGTNTQDYGTALDAAQDEAGGHRLTGNAMSVVSGRVSYTFGFEGPA
LTVDTACSSSLVALHMAAQALRQGECSLAVAGGVTVMATPSSFVEFARQRGLAPDGRC
KPFAAAADGTGWSEGVGLLLVERLSDARRNGHQVLAVVRGSAVNQDGASNGLSAPSGP
SQQRVIRQALANARVAASEVDAVEAHGTGTTLGDPIEAQALLATYGQERPLLLGAVKS
NLGHTQAAAGVAGVMKMVLAMRHGMLPRTLHVDEPTGHVDWTAGAVELLTEHTDWPET
GHPRRAAVSAFGISGTNAHVVLELPAAEQPLVEQPSAAEPDAPATAPDRTPTASDGTA
PLLLSAKSESALRAQAARLHSHLERDPALRLTDAAYTLMTHRTAFAHRAAVRAADHEA
ALRALTALAAGEADPAVDTGTAHTGRDAVLFSGQGSQRIGMGRELSGRYPVFAEAFDT
VCAALDEHLDRPLRDVVRGEDEELLNRTVYAQAGLFAIEVALFRLVESWGVRPHYVAG
HSVGEIAAAHVAGVFSLADACALVAARGRLMQALPAGGAMAAIRATEDEVLPHLADSV
SIAAVNGPSSVVVSGAEHAVLSIAAHFEGAGRKTTRLRVSHAFHSPLMDPMLADFRAV
AEGLTYGEPELAVVSNVTGQLATPDQLRTPEYWVTHVRAAVRFADGIRALGAEGVTRF
LELGPDGVLSALARESAPDDAVCTPVLRKDRSEAATLLAALTHLHVHGTEIDWTAFLA
GRDAHAVDLPTYAFQHQRFWPTPDHTRTGDLGAVGLEATGHPLLSAAVELPDGEGLLF
TTRLSLQTHPWLAGHVVMGSVLLPGTAFAELALRAADEVGCDRVDELTLAAPLVLPEH
GGVQLQLRVGPADASGRRTLTARSRAEGDGDRPWVQHATGVLAEGESTPEPGYDFHTE
SWPPADAAPVELSGLYPDFAAHGFDYGPHFQGLRTAWRRGDEVFAEVALPAEAEGEAS
AYGLHPALLDAALHVVAFNGVDRGVVPFSWESVALHATGASAVRIRVVRHSGDTVSVD
VADTTGEPVASIGTLVLRAVSADQLAGGADPAVRDALFRVQWNPVRLPPAGAAVTVAT
LGSLAGAPFDGYPDLASLARSGRVAGAVLVPVEAGAGEVVADDVVGATHATAARALDL
ARSWLADDRFAASRLVFVTRGAVSGADLAGAAVWGLVRSALSEHPGRFGLVDLDDDAE
LALVPRVLASDEPQLLVRGGEVLAARLARAQSSHAVTWDPSGTVLVTGGTGGLGRVMA
RHLVVEHGVRNLLLVSRRGPAAEGAEELVTELRHSGAEVAVEACDVTDAAAVADLVAR
HRISAVVHTAGVLDDGVVESLTPERLSAVLRPKVDAAWNLHEATRDLDLDAFVVFSSV
AGTIGSPGQANYAAGNAFLDALAHHRRAAGLPAASLAWGPWSRDGGMTGTLTDVDSSA
SPGRHARTHPRTGRGLFDAALAAGDAHLLPVRFDWASLRAQGEVPPLLRGLIRTRARR
SAVGGSAAAAGLVGRLSGRGTVERREVLLDLVRAQIAVVLGHANPETIESTRVFQDLG
FDSLTAVELRNRLNNATGLRLSATAVFDYPTADALVDFLLDELFGAQEEAELPAPVPS
PAGAADDPVVIVGMSCRYPGGVGSPEDLWRLVSEGVDAVSDFPTDRGWDVESLYSPDP
EALGTSYTRSGGFLHEAAEFDPDFFGMSPREALATDAQQRLLLETTWEAIERTGIDPA
SLRGSRTGVFAGVMYTDYGDLLVGDQFEGYRSNGSAASIASGRVSYTFGFEGPAVTVD
TACSSSLVALHWAAQSLRSGECSLAVAGGVTVMSTPTTFVEFSRQRGLSADGRCKAFA
DAADGVGWGEGVGMLVLERLSDARRNGHRVLAVVRGSAVNQDGASNGLTAPNGPAQQR
VIRQALASAGLSAADVDAVEAHGTGTTLGDPIEAQALLATYGQERPEDRPLLLGSVKS
NIGHAQAASGVAGVIKMVLAMRHGVLPRTLHVDEPSSHVDWSAGAVELLTSEAEWPQG
EGPRRAGVSSFGVSGTNAHVILEQPGPDAADAAPDATVTDPGALAWVLSARNEAALRC
QAARLLSLVAGSDALCARDIGHSLVTGRSSFAHRAVVWGQDRDALVRALSALAVGEAD
AGLAEGASGAGRTAFLFSGQGSQRLGMGWELYARYPVFADAFDAVCAALDEHLERPLR
DVVWGEDAELLNQTAYAQAGLFAIEVALYRLAESWGMRPDFVAGHSIGEVAAAHVSGV
FSLPDACALVAARGRLMQQLPSGGAMMAIRATEDEVLPHLAEGVSLAAVNGPSSVVIS
GAEDAVLAIAAHFAGEGRKTTRLRVSHAFHSPLMEPMLEEFRAVVTRLSFGTPTIPVV
SNLTGRLAEPEQLAHADYWVRHVREAVRFADGIQALRAEGVTRFLELGPDGVLSAMAR
ESASDDAVLAPVLRRDRPEETALLGALAQLYVRGAHVDWTVPFAGSGARWADLPTYAF
QHERFWPSGGVARPGDVRSAGLGSAGHPLLGAAVELAGSGGLLFTGRLSVSSHPWLAD
HVVLGSVLVPGTALVELVLRAADEAGCDLLEELTLAAPLVLPASGAAVQVQVAVGEPD
EAGRRPVSVHAREGEGPWTLHASGAVTSGAEVPPFDATVWPPKGAEPVDVADCYDVLA
DAGLTYGPAFHGLQAAWKLGGDVYAEAKLPESTDGDAYGLHPALFDAALHASALGGAE
AGGVPFSWAGVSLHATGASHLRVRIREAGGALSVAIADTSGAPVASVESLVIRPLSAG
QVQAADRDALFKADWVPVPLTDERVEPGTGPEGEPLRTYADLDSLEGAAVPGTVLVAP
PSGAAGTVESVHAATVWALEMVQAWLADDRFATSRLVFVTRGAAFGADLAAAAVRGLV
RSAQSENPGRFGLVDMDGDADTTVPAQALATDEPELLVRGGEVLAARLVRAQSSHTVT
WDPSGTVLITGGTGGLGRSVARHLVSEHGVRSLLLVSRRGPAAEGAGELVAELRGSGA
EVVIEACDVTDAVAVADLVARHRISAVVHTAGVLDDGVVESLTPERLAVVLRPKVDAA
WNLHEATRGLDLDAFVVFSSVAGTFGSAGQANYAAGNAFLDALAYHRRAVGLPAVSLA
WGPWSQDGGMTGTLSDADVQRIARQGMPPLTVEEGLALFDAALGSAEPMALPVRLDLA
ALRAQGEPQPLLRGLIRTRTRRSGAAAASGIAQRLAGLSTAERREALLDVVRAQIATV
LGHAGPETIAPDRAFQDLGLDSLTAIELRNLLGKATGLRLPATTVFDYPTVDALAAHL
LDELFGAETGTATETPLPVPGLPSLADDPVVIVGMSCRFPGGVASPEDLWRLVADGVD
AVSAFPTDRGWEIDDTYDPEREGAIATRSGGFLHDAAEFDPEFFGMSPREALTTDAQQ
RLLLETTWEALERAGMDPATLRGSRTGVFAGVMYHDYSTLLSGREFEGYQGSGSAGSV
ASGRVSYTFGFEGPAVTVDTACSSSLVALHLAAQSLRSGECSLALAGGVTVMSTPLTF
VEFSRQGGLSADGRCKAFADAADGVGWAEGAGILVLERLSDARRNGHRILATVRGSAV
NQDGASNGLTAPNGPAQQRVIRQALASAGLSAADVDAVEAHGTGTTLGDPIEAQALLA
TYGQERPEDRPLLLGSVKSNIGHAQAASGVAGVIKMVLAMRHGVLPRTLHVDEPSSHV
DWSAGAVELLTSEAEWPQGEGPRRAGVSSFGISGTNAHVILEQPEPVAAETESITPDT
APDAAEDEAADSGTPVPALLSGRSASALRAQAARLLSRLDGDPGPRITDVAYSLATGR
SAFPHRAVILAANRADLLHSLSALAEGHTEAPAVVAQDRARSGKLAFLFSGQGSQRLG
MGRELYGRYPAFAEALDAVCAALDAHLDRPLRDVIWGEDAELLNRTGYAQTGLFAIEV
ALFRLLESWGVRPDHLLGHSIGEIAAAHVAGVLSLPDACALVAARGRLMQQLPSGGAM
MAIRATEDEVLPHLAEGVSLAAVNGPSSVVVSGAEDEVLALAAHFEEEGRKTTRLRVS
HAFHSPLMEPMLADFRAVADGMTYAAPRIPVVSNVTGRPATAEELCCAEYWVGHVREA
VRFADGVGALREQGVTTFLELGPDGSLSALAAESAADDSVLAPVLRKNRPEAPALLTA
LARLHAQGTPVDWSAAFAGTGARWVDLPTYAFQHERFWPSGGAARAGDVRSAGLGSAG
HPLLGAAVELAGSGGRLLTGRLSLSSHPWLADHVVLGSVLVPGTALMELVLRAADEVD
CAAVDELTLAAPLVLPASGAAIQVQVWVGEPDEAGRRPVSVHAREGEGPWTLHADGAL
APAAETVPFDTAIWPPQGAEHLDAAGCYERFADAGFAYGPVFQGLRAAWKLGEDIYAE
VALPEGTDGNAYGLHPALFDAALHAALLGGEGTDEAAVPFSWNGVTLHATGASRVRVR
IRPTEGGTSIALVDTAGAPVASVRSLTARPITAGQLQTGDRDSLFQVDWTTLHLTDER
ANSLALLGKDTEGILDTLSLQPHADLDDLAATGVHDTVLAPLPTRTAGTVESVHAATT
GALALIRSWLADDRFAASRLVFVTRGAVSGTDLAGASVWGLVRSALLEHPGRFGLVDV
DVDQDAEVPLVPRALASDEPQVLVRGGEVLAARLVRAQSSDTVTWDPSGTVLITGGTG
GLGRSVARHLVSEHGVRSLLLVSRRGPAAEGVDALVAELAECGAQVTVEACDVTDAVA
VADLVARHRISAVVHTAGVLDDGVVESLTPERLSAVLRPKVDAAWNLHEATRGLDLDA
FVVFSSVAGTFGSAGQANYAAGNAFLDALAYHRRAVGLPAVSLAWGPWSQDGGMTGTL
SDADVQRIARQGMPPLTVEEGLALFDAALGSAEPMALPVRLDLAALRAQGEPQPLLRG
LIRTPGRRTAAAATEGDTAAAFAGRLTGLSAAEGREVVLGAVRSQIAGVLGHAEATEI
DQDRAFLDLGFDSLTAVELRNRLGAVTGIRLPATLLFDYPTPAELVAHLHARIAPEPT
VGPEALLGELERMEKSFGGLDLTEEMHEQIAGRLEVLRAKWDALRDTAAAAGHDGSPS
DEDFDFESASDDEVFDLLDNELGLS"

BASE COUNT 2611 a 7486 c 7397 g 2900 t
ORIGIN

Query Match 32.7%; Score 302.2; DB 6; Length 20394;
Best Local Similarity 60.9%; Pred. No. 1.1e-23;
Matches 572; Conservative 0; Mismatches 343; Indels 24; Gaps 4;

```
Qy          1 ACCGCATTCATCTGCTCCGGACAGGGCACCCAACGCCCCGGCATGGCCCACGGCCTCTAC 60
              | ||| ||| ||| | |||| |||||| |||| || || |||||||  | ||  ||  ||
Db       1687 AGCGCGTTCCTCTTCGCCGGGCAGGGCTCCCAGCGACCGGGCATGGGGCGCGAACTGCAC 1746

Qy         61 CACACCCACCCCGTCTTCGCCGCCGCACTCAACGACATCTGCACCCACCTCGACCCCCAC 120
                | | |||||||| |||||| | ||  || |||    | ||| || | || ||||| |||
Db       1747 GCCGCGCACCCCGTGTTCGCCACGGCCTTCGACGCGGTGTGCGCCGAACTGGACCCACAC 1806

Qy        121 CTCGACCACCCCCTCCTCCCCCTCCTCACCCAAAACGACAACGACAACGAGGACGCGGCC 180
              || ||||  || || | |  | || ||          || | | ||   |  | || ||||
Db       1807 CTGGACCGGCCGCTGCGCGACATCGTC------TTCGCCGAGGAAGGCAGCGCCGAGGCC 1860

Qy        181 GCACTGCTCCAGCAGACCCGCTACGCCCAGCCCGCCCTCTTCGCCTTCCAGGTCGCCCTC 240
              || |||||| | ||||||  |||| ||||| |||| |||||||||| | |  ||||||
Db       1861 GCCCTGCTCGACCAGACCGCCTACACCCAGGCCGCGCTCTTCGCCCTGGAAACCGCCCTG 1920

Qy        241 CACCGCCTCCTCACCGACGGCTACCACATCACCCCCCACTACTACGCCGGACACTCCCTC 300
                |||  ||| |   |||   ||    | |  | ||||  | |  ||||||||||||| ||
Db       1921 TTCCGGCTCGT---CGAATCCTGGGGCGTGGCACCCCGGTTCGTCGCCGGACACTCCATC 1977

Qy        301 GGCGAAATCACCGCCGCCCACCTCGCCGGCATCCTCACCCTCACCGACGCCACCACCCTC 360
              |||||  | ||||||||||||  ||  ||| | || ||||||  |||||||  |  ||
Db       1978 GGCGAGCTGACCGCCGCCCACGTCAGTGGCGTGCTGACCCTCCACGACGCCGCACGGCTG 2037

Qy        361 ATCACCCAACGCGCCACCCTCATGCAAACCATGCCCCC---CGGCACCATGACCACCCTC 417
              || ||   |||| |||||||||||||  |  |||||  |   |||| |||||    |  ||
Db       2038 GTCGCCGCGCGCGGCACCCTCATGCAGGCGCTGCCCGCAGGCGGCGCCATGGTGGCGGTC 2097

Qy        418 CACACCACCCCCCACCACATCACCCACCACCTCACCGCCCACGAAAACGACCTCGCCATC 477
              ||  |||||   || | |||  | | |  ||| ||| ||||||  || || ||||| ||
Db       2098 CAGGCCACCGAGGACGAGATCCGCGAGCGTCTCGCCGGCCACGAGGACCACGTCGCCCTC 2157

Qy        478 GCCGCCATCAACACCCCCACCTCCCTCGTCATCAGCGGCACCCCCCCACACCGTCCAACAC 537
```

```
                 || |||  ||||  |||  |||  ||||||    |||  |   ||  |||||    |
Db       2158 GCGGCCGCCAACGGGCCCGATTCCACCGTCATTTCGGGCGACGAACAGGCCGTCACCGAG 2217

Qy        538 ATCACCACCCTCTGCCAACAACAAGGCATCAAAACCAAAACCCTCCCCACCAACCACGCC 597
                 ||| |  ||| |||  |   ||| |||  |   |||||    || |    || |||||||
Db       2218 ATCGCGGCCCACTGGGAGGCACAGGGCCGCCGCACCAAGCGGCTGCGGGTCAGCCACGCC 2277

Qy        598 TTCCACTCCCCCCACACCAACCCCATCCTCAACCAACTCCACCAGCACACCCAAACCCTC 657
                 ||||||||||| ||||   ||  ||| ||  |  |  |||  |    | |||      ||
Db       2278 TTCCACTCCCCGCACATGGACGACATGCTGGAGGACTTCCGGCGCGTCGCCCGCGGTCTG 2337

Qy        658 ACCTACCACCCACCCCACACCCCC------------CTCATCACCGCCAACACCCCACCC 705
                 |||| |||| | |||| || ||||           | | |  |||   | || |
Db       2338 ACCTTCCACGCCCCCCGCATCCCCGTGGTGTCCACGGTGACCGGCGCGCTCGCCACCGAA 2397

Qy        706 GACCAACTCCTCACCCCCCACTACTGGACCCAACAAGCCCGCAACACCGTCGACTACGCC 765
                 ||| |||| | | | ||| ||||||||    |  || | |||| | ||||||  || |
Db       2398 GACGAACTGCGCTCGCCCGACTACTGGGTGCGGCAGGTCCGCGAAACCGTCCGCTTCTGT 2457

Qy        766 ACCACCACCCAAACCCTCCACCAACACGGCGTCACCACCTACATCGAACTCGGACCCGAC 825
                  |  ||   |  |||||  |     | ||||||||||||| | |  ||  ||||  ||| |
Db       2458 GCTGCCGTGCGCACCCTTGAGGCCGAGGGCGTCACCACCTTCGTGGAGATCGGCACCGGC 2517

Qy        826 AACACCCTCACCACCCTCACCCACCACAACCTCCCCAACCCCCCCACCACCACCCTCACC 885
                  |  ||||||| || |   |||  ||   ||  ||| || |       ||   |||  |
Db       2518 GGCGTCCTCACCCCCATGGTCCAGGACTGTCTGACCACCCTCGAAGAGCCCGTTCTCGTC 2577

Qy        886 CTCACCCACCCCCACCACCACCCCCAAACCCACCTCCTC 924
                 |     |  | |  |  || |||| |||||  |   |||
Db       2578 CCTCTGCTGCGCACCGGCCGCCCCGAAACCGTCGCTCTC 2616


RESULT 13
SNA278573/c
LOCUS       SNA278573              84985 bp    DNA     linear   BCT 17-MAY-2001
DEFINITION  Streptomyces natalensis pimaricin biosynthetic gene cluster.
ACCESSION   AJ278573
VERSION     AJ278573.1  GI:12055067
KEYWORDS    ABC transporter; cholesterol oxidase; cytochrome P450 monoxygenase;
            ferredoxin; glucosyl transferase; macrolide efflux pump; mycosamine
            dehyralase; orfX; pimA gene; pimB gene; pimC gene; pimD gene; pimE
            gene; pimF gene; pimG gene; pimH gene; pimI gene; pimJ gene; pimK
            gene; pimS0 gene; pimS1 gene; pimS2 gene; pimS3 gene; pimS4 gene;
            polyketide synthase; sensory transduction protein; thioesterase.
SOURCE      Streptomyces natalensis.
  ORGANISM  Streptomyces natalensis
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1  (bases 1 to 84985)
  AUTHORS   Aparicio,J.F., Fouces,R., Mendes,M.V., Olivera,N. and Martin,J.F.
  TITLE     A complex multienzyme system encoded by five polyketide synthase
            genes is involved in the biosynthesis of the 26-membered polyene
            macrolide pimaricin in Streptomyces natalensis
  JOURNAL   Chem. Biol. 7 (11), 895-905 (2000)
  MEDLINE   20547809
   PUBMED   11094342
REFERENCE   2  (bases 1 to 84985)
  AUTHORS   Aparicio,J.F.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-JUN-2000) Aparicio J.F., Molecular Genetics,
            Institute Of Biotechnology - Inbiotec, AV. REAL 1 - LEON, 24006,
            SPAIN
FEATURES             Location/Qualifiers
     source          1. .84985
                     /organism="Streptomyces natalensis"
                     /db_xref="taxon:68242"
     CDS             <1. .2086
                     /note="ORFX"
                     /codon_start=2
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="CAC20917.1"
                     /db_xref="GI:12055068"
                     /translation="TRPDAFREATEPYGALLAPLMPERSADRVEGTKWTSDWSQARFL
                     AHDAVCEVLLTLAGQHPLVLLLEDLQWADTASLDLLRLLSPRCQGHPLGIVLTARKFE
                     TESDATLRRIVSDVLRGPRTATLRLCGLPRQAVGALVEAHAGSGVDPEVVEVLHRRSE
                     GNPYFVMQLLSLVGDARRLHRPDAVDALLAHIPTGVREALNQRFAALPEPVLRVLRLC
                     AVIGAEVDTDLLRRTATCDASVNAGLESAIRAGLLGEDPHHPGQLHFAHALVQETLFD
                     ELPRADRRRLHAKVADALYARRRGHVGDEEIERVAHHSWHAKSALPAAEVLPRLLLAA
                     EHAEQAMAYEQVEIWLRRAVHLVGFLPPDDPATVRLEQKLHIQLGQVLATTRGYGHPE
                     AETALTRGRALSTVTHAPEDPSVLWALCAASLVTGRYDASRQLSGLLRDLSWQTREPV
                     AVLGAAYGEGIVLHVRGMLPQALAELEHGVERADHFANEGHALARTFQHDPRVSCRSY
                     NTFTHWLLGHRQIADARRRELLSLTEYESRPSDRSFALYVDGVVAAWEGDTRTALASS
                     DEGVRVAGDHGLLYWKAMLGVVKGWGLVHAGDRDEGLALMHTSLAELGQSRTYLRHPL
                     HLGLLGQAQHHNGQIEEARTTFRMLLAAVEQRRERVYLHPALPATPLLHELLGHGADE
                     ALADRDPGRGQV"
gene                 2190. .3566
                     /gene="pimK"
CDS                  2190. .3566
                     /gene="pimK"
                     /function="putative glucosyl transferase"
                     /codon_start=1
                     /transl_table=11
                     /product="PimK protein"
                     /protein_id="CAC20918.1"
                     /db_xref="GI:12055069"
                     /translation="MESARRPILFVSLPESGLLNPLLVLAGELSRQGVEDLWFATDEP
                     RRNDVKRIAEGSPVEFASLGEVDSEMSAVTWSDEVYREVTQPSRFKAHRAVVRHTYRP
                     GLQAEKFRRLQAVIDEVQPALMVIDCISGFAVDAAIARNIPYVLSVPFLPSNVLTAHT
                     HFAKSYTPRGFPVPHTGLSRRMTLAQRVANELFKLRTFAMFLNPRLGKVLAEDNRRRN
                     ELGLPKASFMARIEHADLVLCNSLAELDYPFDIPEKMRLVGAMVPPLPEAPDDQDLSR
                     WLDAQSSVVYVGLGTITRLTREQVGSMVEVARRLEDRHQVLWKLPSEQQHLLPPRESL
                     PGNLRVESWVPSQMDVLAHPHVKVFFTHGGGNGFNEGMYFGKPLVVRPLWVDCYDQAV
                     RGQDFGLSLTLDRPQTIDVNDVVDKLTRVLGTPSFYEKAERRAALMRSAGGRETAAGL
                     VLSLPALA"
gene                 complement(3662. .9736)
                     /gene="pimS4"
CDS                  complement(3662. .9736)
                     /gene="pimS4"
                     /function="polyketide synthase"
                     /codon_start=1
                     /transl_table=11
                     /product="PimS4 protein"
                     /protein_id="CAC20919.1"
                     /db_xref="GI:12055070"
                     /translation="MPNDKPNDKKLVDYLKWVTNDLHQTRQRLREVESAKQEPIAIVG
                     MSCRLPGGVRSPEEFWQLLKDGRDGISTFPTDRGWDLDLLAGDGSGSSATAEGGFIDA
                     ASFDAGFFGISPREAVTMDPQQRILLESSWEALERAGIDPVSLRGTETGVFVGTSGID
                     YINTVMNSREDIEGHGSTGLIASILSGRVSYTFGLEGPAVTVDTACSSSLVSLHLAAQ
                     ALRVGECSLALAGGVTAISTPMSFVGFSRQGGLAPDGRCKAFSDTADGTAWSEGVGVL
                     VVERLSDARRNGHPVLAVVRGSAVNQDGASNGLTAPSGPSQQRVIRQALAAAGLSASD
                     VDVVEAHGTGTALGDPIEAQALLATYGRDRDPERPLLLGSVKSNLGHTQAAAGAAGLI
                     KMILAMRHGVLPRTLHVEAPSSHVDWSSGALQLLTEPQEWPKADRPWRAGVSSFAISG
                     TNAHIILEQPPAAEEPEDEEGPHTSPSAVPWLVSAKSEEALTAQLDRISALTDRSPLD
                     LGLSLATTRAHLEHRAVLLSEGGETVEVARGHAEHTAGKLAVLFSGQGAQRVGMGREL
                     YGRFPVFAEALDGVLAHFDAGLRDVMFGEAEGLDETGFTQPALFAIEVALFRLSESLG
                     VRPDFVAGHSIGEIAAAHVAGVFSLADACALVAARARLMQALPAGGAMVAVQAAESEV
                     AARLVDGVSIAAVNGPEAVVIAGEEREVLRIAEEFTGLGRKTRRLPVSHAFHSPLMEP
                     MLEDFRRVAEGLSYEAPRIALVSNLTGELATEDLVRSAEYWVRHVRETVRFADGVSAL
                     AAEGASVFLELGPDGVLTAMAQHTLDGTATTVPALRKDRSEETALLTALAQLHVAGVG
                     VDWSGIFAGTGARRVDLPTYPFQHERYWPRPAVLSGDVVGAGLRSAEHPLLGATLTLA
                     ESGEVVLTGRLSLQAYPWLTDHTMGGSVLFPATGFLELAMRAADQVGCDRVDEFVLMT
                     PLVLAEESGTQVQVLVGPADEAGTRTISIHARPDGDPDQPWMQHAGGTLTSGERIVDF
                     GSTVWPPQGAEAVDLEGFYDTTGYGPAFQGLRSVWQCEDGAYVEVALPDQVAEDAGAF
                     GLHPVLLDAVLQSRRMAGIGAAGDTFLPFAWRGVSLHAGGASVLRARVIKTGEDTLSL
                     VAVDAEGAPVLSVEELVLRTQLTHTDAPLTARRGAERDALLSLEWVSGPEVPPAGVRC
                     VSLGADELGVGASVATLADLSGDEDFVVVPVSGVGGDVPAAAHELTARTLGLLQEWSG
                     REGTSASRLVFVTRNAVSADGGESVRDVAAAAVWGLVRSAQSESPGRIALVDLEADSG
                     DLAEVLAGLPGLLAAGDAQFVVREGAVRVGRLVESNEGAGVSPREWDADGTVLVTGGT
                     GGLGRELARHLVVERGVRHLLLVSRSGPEASGVGELRDELAAHGAEIAVRACDVADRA
                     AVEAVVGAVDAAHPLTAVIHTAGVLDDGVIASLTPERLSAVLRPKVDAAWNLHEATKD
                     LDLAAFVLYSSISGVLGSLGQGNYAAGNVFLDALAGHRAGLGLPAQSLAWGAWTQGGG
                     MTATLSDADMQRMASYGTAPLSPERGLALFDAATAGSATHLVAVGRVSGPVRAQGPVP
                     SMLRGLVKGARRTAVNTTGETAVSLADRLRAARPEERTQLMTDVVRTEAAAVLGHASS
                     QNVDARREFYELGFDSLTSVELRNRLSTVTGLRLSATVVFDSKNPTDLAARLYADLAA
                     QSALDGSASQAGFLPVAAPETDSLERMFLDALDSGKVPEAQRMLSALGALRPSFENTA
```

```
                ELEDLPLPATLAEGPGTPRLICVSTPTANGGVHEYARLAASFRGERHVSALPLVGFAT
                GERLPATAETAVRVIAESALRASDGNPPVLVGHSSAGAFAYLAAALLENTWGIRPEAV
                VLLDTLSLRHEQNESIDYAGLMRRHFMVDEVSPVRMTNSRLSAMARWMGMLNQLEVRH
                TTVPVLIIRATKETFGIGTDTGIYGEDHGSPVDVRSVDADHFSMVRDDAPETARIVRE
                WLDSLGNA"
gene            complement(9772. .15198)
                /gene="pimS3"
CDS             complement(9772. .15198)
                /gene="pimS3"
                /function="polyketide synthase"
                /codon_start=1
                /transl_table=11
                /product="PimS3 protein"
                /protein_id="CAC20920.1"
                /db_xref="GI:12055071"
                /translation="MTTSSETVVAALRASLKEAENLRRQNRKLVAAATEPIAIVAMAC
                RYPGGVQSPEDLWELLTAGKDAIGAFPTDRGWDLDALRDAGVDRRGNAVSQEGGFLDG
                VGDFDADFFGISPRESVTMDPQQRLLLETSWEAIERAGIDPTTLRGSRTGVFIGTNGQ
                DYAYLLVRSLADATGDIGTGIAASATSGRLSYTLGLEGPAVTVDTACSSSLVALHTAS
                HALRAGECSLALVGGVNVMSSPGSLMEFSRQGGLAADGRCKAFSDTADGTGWSEGVGV
                LVVERLSDARRNGHPVLAVVRGSAVNQDGASNGFTAPSGPSQQRVIRQALAAAGLSAS
                DVDVVEAHGTGTPLGDPIEAQALLATYGQDRDPERPLLLGSVKSNLGHTQAAAGAAGL
                IKMILAMRHGVLPKTLYADVPSSHVDWTSGEVELLTEPQEWPKADRPWRAGVSSFGIS
                GTNAHVIVEQAEPDEEPVAEPGLAPAVVPWLVSAKTEEALAAQLARISSRRGTASPVD
                VGFSLASGRSAFEHRAVLLAQCPDGELSEAARGRAETSRPLAVLFSGQGAQRVGMGRE
                LYVRFPVFAEALDEVLAHFDAGLRDVMFGDAEGLDETGFTQPALFAIEVALFRLAESL
                GVRPDFVGGHSIGEIAAAHVAGVFSLTDACALVAARARLMQALPAGGAMVAVQASESE
                VAARLVDGVSIAAVNGPEAVVLAGDETAVLRIAEELAAGGRKTQRLPVSHAFHSPLME
                PMLEDFRRVAEGLSYGAPQIPLVSNVTGELAVEELVCSPDYWVRHVRETVRFADGVSA
                LEAEGASAFLELGPDGVLTAMARHTLDETKVAVPVLRKDRPEETALLTALAHLHVAGI
                GVDWSGIFAGTGARRVDLPTYPFQHERYWPRPAVLSGDVAGAGLRSAEHPLLGATLTL
                AESGEVVLTGRLSLQTHPWLTDHAIGGNVFFPGTGFLELAMHAADQAGCDRVEDLVLM
                TPLVLAEENGTQVQVLVGPADEAGTRTISIHARPEDDPDQPWMQHAGGTLTSGERVVD
                FGSTVWPPQDAVAVDLEGFYEGTGYGPAFQGLRSVWSCGDGAYAEVALPDHVAEDAGA
                FGLHPALLDAVLQSHRIAGVGSADDQFLPFAWRGVSLHAGGASVLRVRVVKTGDDSMS
                IAAVDAEGAPVLSAEAMVMRAYRDPGTGQAAARHGAERDELLSLEWVSGPEVPPAGVR
                CVSLGADELGVGASVASLADLSGDEDFVVLPVSGVGGDVPAAAHELTARTLGLLQEWS
                GREGSSGSRLVFVTRNAVSADEGESVRDVAAAAVWGLVRSAQSESPGRIALVDLEADS
                GDLAEVLAGLPGLLAAGDAQFVVREGAVRVGRLVESNEGAGVSPREWDADGTVLVTGG
                TGGLGRELARHLVVERGVRHLLLVSRSGPEAPGVSGLSDELAAHGAEVTVRACDVADR
                AAVHALVAAVPAEHPLTAVIHTAGVLDDGLVTSLTPERLSTVLRPKVDAAWNLHEATK
                DLDLAAFVLYSSISGVLGAPGQANYAAGNVFLDALAWHRMGQGLPAQSLAWGAWTQDG
                GMTATLSDTDMQRMASYGTAPLSRQRGLALFDAACARSAAHLVAVGRVGGPVRASAPV
                PSVLRGLVKGARRTAVNTTGETAVSLADRLRAARPEERPRLMTDLVRTEAAAVLGHAS
                AKNVDARREFNELGFDSLTSVELRNRLSALTGLRLSAALVFDYPNPTALADHFVSELV
                EEAATGPALLAELERLDSALAAGAPDARTRAAVTARLAQILDDWRDAPAEEDGAEVAE
                RLEAASTDEIFAFIDNELGRHSDR"
gene            complement(15210. .43733)
                /gene="pimS2"
CDS             complement(15210. .43733)
                /gene="pimS2"
                /function="polyketide synthase"
                /codon_start=1
                /transl_table=11
                /product="PimS2 protein"
                /protein_id="CAC20921.1"
                /db_xref="GI:12055072"
                /translation="MDQEQKLRDYLKRASADLRRSRQRVGELEAAAREPIAIVGMSCR
                FPGGVRSPEELWSLLAADGDGITKSPADRGWELLAEDAGEFSGGFLHDAAEFDASFFG
                ISPREALAMDPQQRLLLESSWEAFERAGIDPAAVRGTRTGMFVGAMPQEYRVGPDDDV
                QGFALTGTTTSVISGRLAYFFGAVGPAVTVDTACSSSLVALHLAAHSLRQGECSLALA
                AGVTVMSSPTTFVEFNRQGGLSADGRCRSFADSADGTGWSEGVGVLVLERLSEARRNG
                HEILAVIRGSAVNQDGASNGLTAPNGPSQRRVIEQALVSARLSADEVDVVEAHGTGTT
                LGDPVEAQALLATYGQGRDDDRPLLLGSVKSNLGHTQAAAGMAGVIKMVLAMRHGVLP
                RTLHVDAPSSHVDWTQGAVRLLTEHVPWPQGEQPRRAGVSSFGLSGTNAHTILEEAPA
                APSPEAGQEDTAPVAEPLAAGAVPWLVSGRTREALRAQAARLLEHLTSRPGLAPLDVA
                YSLATARSGLEHRAAFTAAGPDGARAALTALADGTPAPGLVQDTARTRSKLAFLFAGQ
                GSQRPGMGRELAARFPVFATALDEVLGHFDEGLERPLKDVLFAAEGTPEAALLDQTGY
                AQPALFAIEVALYRLAESFGITPDFLAGHSIGEIAAAHVAGVFTLADAAALVLARGRL
                MQALPEGGAMVSLEATEEEVLPLLEDRPDRLSIAAVNGPAAVVIAGEADDVTAVAEHF
                DALGRRTKRLRVSHAFHSPLMEPMLAEFRTAVSRLTPQAPVIPVVSGLTGTLATVEQL
                TSPDYWLDHARHAVRFADALRLARRARRRHLPRMGPDGVLSAMAQSCCRGHGHRRRRR
                LRHARRVRPGRPEAETLTAALAGLHTRGVAVQWEPYFQGTGARRADLPTYAFQRRRYW
```

```
                PKSLPATGGDVRAAGLGAAHHPLLTAAVSVANSDGLLLTGRLSRRTHPWLADHAVRGT
                VLLPGTAFLELAVRAGDEAGCGRVEELTLAAPLVLPEEGGVQVQVWVGSPDASGRRAV
                SVHSRPDGPEELPWTQHAAGTLVAGEHHAGFDATVWPPADATPLDLDGFYARMADTGF
                GYGPLFQGLRAAWRSGDDVYAEVALPDSDGAASTESSAAFGLHPALLDAALHAAAFVD
                LGEDARGGLPFSWQDVTLHASGASTVRVKLTPDGDDAVALAVADTTGAPVASIGSLVL
                RTLPDEQIDAAHSLVRDALFELRWSEARQTAHEAPASAAVLGPDLFGLERALAPTDLT
                VTAVPGPDAAPHAPDVVLAPVAGTTEDDVVTSVHARTAHVLERLRTWLTDERYAASRL
                VPVTRGALATDARQASDPVAAAVWGLVRAAQSEHPGRFGLLDLDPGTTEPATEPLLRA
                LALSSDEPQTAVRGTAVLTARLARAQAPQAATEWDPEGTVVLTGATGGLGRVLARHLV
                AERGVRHLLLVSRRGADADGAGELMAELAAHGAQVTLAACDLADRAAVRRLLSGVPAE
                HPVTAVVHSAGSLDDGVIASLTPERISAVLRPKADAVWHLHEATRDLDLAAFVVFSSL
                SGTVGAAGQGNYAAANAFLDALAQLRRAQGLPGLSLAWGPWAPTAGMTGGLTDDDLDR
                LARMGTPALTEEQGIALFDAATATDAAVLLPTRIDLSVLRVQSEIPPLWRGLIRTPAR
                RSVVSASETAVTLVQQLSRLAETDRREVVLDLVAGQVAAVLGHAGTADIDPRRPLREL
                GFDSLTAVELRNRLSAATGLRTAATVIFDHPTVDALAAHLLDGLMGSEAAAAADAPRT

  Query Match           32.7%;  Score 302.2;  DB 1;  Length 84985;
  Best Local Similarity 60.9%;  Pred. No. 8.2e-24;
  Matches  572;  Conservative    0;  Mismatches  343;  Indels   24;  Gaps    4;

Qy          1 ACCGCATTCATCTGCTCCGGACAGGGCACCCAACGCCCCGGCATGGCCCACGGCCTCTAC 60
              | ||| ||| ||| | |||| |||||| |||| || || |||||||  | ||  ||  ||
Db      79466 AGCGCGTTCCTCTTCGCCGGGCAGGGCTCCCAGCGACCGGGCATGGGGCGCGAACTGCAC 79407

Qy         61 CACACCCACCCCGTCTTCGCCGCCGCACTCAACGACATCTGCACCCACCTCGACCCCCAC 120
              | | |||||||| |||||| | ||  || ||| |  ||| || | || ||||| |||
Db      79406 GCCGCGCACCCCGTGTTCGCCACGGCCTTCGACGCGGTGTGCGCCGAACTGGACCCACAC 79347

Qy        121 CTCGACCACCCCCTCCTCCCCCTCCTCACCCAAAACGACAACGACAACGAGGACGCGGCC 180
              || ||||  || || | |  | || ||        || | | ||   |   | || ||||
Db      79346 CTGGACCGGCCGCTGCGCGACATCGTC------TTCGCCGAGGAAGGCAGCGCCGAGGCC 79293

Qy        181 GCACTGCTCCAGCAGACCCGCTACGCCCAGCCCGCCCTCTTCGCCTTCCAGGTCGCCCTC 240
              || |||||  | ||||||  |||| |||||  |||| |||||||||  |  |   ||||||
Db      79292 GCCCTGCTCGACCAGACCGCCTACACCCAGGCCGCGCTCTTCGCCCTGGAAACCGCCCTG 79233

Qy        241 CACCGCCTCCTCACCGACGGCTACCACATCACCCCCCACTACTACGCCGGACACTCCCTC 300
              |||  |||  |   |||  ||   | |  | ||||  | |  ||||||||||||| ||
Db      79232 TTCCGGCTCGT---CGAATCCTGGGGCGTGGCACCCCGGTTCGTCGCCGGACACTCCATC 79176

Qy        301 GGCGAAATCACCGCCGCCCACCTCGCCGGCATCCTCACCCTCACCGACGCCACCACCCTC 360
              |||||  | ||||||||||||| ||   ||| | || ||||||  |||||||  |   ||
Db      79175 GGCGAGCTGACCGCCGCCCACGTCAGTGGCGTGCTGACCCTCCACGACGCCGCACGGCTG 79116

Qy        361 ATCACCCAACGCGCCACCCTCATGCAAACCATGCCCCC---CGGCACCATGACCACCCTC 417
              || ||   |||| |||||||||||  |  |||||  |   |||| |||||     |  ||
Db      79115 GTCGCCGCGCGCGGCACCCTCATGCAGGCGCTGCCCGCAGGCGGCGCCATGGTGGCGGTC 79056

Qy        418 CACACCACCCCCCACCACATCACCCACCACCTCACCGCCCACGAAAACGACCTCGCCATC 477
              ||  |||||    || | |||  | | |  ||| ||| ||||||  || || ||||| ||
Db      79055 CAGGCCACCGAGGACGAGATCCGCGAGCGTCTCGCCGGCCACGAGGACCACGTCGCCCTC 78996

Qy        478 GCCGCCATCAACACCCCCACCTCCCTCGTCATCAGCGGCACCCCCCACACCGTCCAACAC 537
              || |||  ||||  |||  |||  ||||||  |||  |  ||  |||||   |
Db      78995 GCGGCCGCCAACGGGCCCGATTCCACCGTCATTTCGGGCGACGAACAGGCCGTCACCGAG 78936

Qy        538 ATCACCACCCTCTGCCAACAACAAGGCATCAAAACCAAAACCCTCCCCACCAACCACGCC 597
              |||  |||  |||  |   ||| |||  |  |||||   || |   || |||||||
Db      78935 ATCGCGGCCCACTGGGAGGCACAGGGCCGCCGCACCAAGCGGCTGCGGGTCAGCCACGCC 78876

Qy        598 TTCCACTCCCCCCACACCAACCCCATCCTCAACCAACTCCACCAGCACACCCAAACCCTC 657
              ||||||||||| ||||   |  ||| || |  |  |||  |    | |||      ||
Db      78875 TTCCACTCCCCGCACATGGACGACATGCTGGAGGACTTCCGGCGCGTCGCCCGCGGTCTG 78816

Qy        658 ACCTACCACCCACCCCACACCCCC------------CTCATCACCGCCAACACCCCACCC 705
              |||| |||| | |||| || ||||                | | |  |||   | || |
Db      78815 ACCTTCCACGCCCCCCGCATCCCCGTGGTGTCCACGGTGACCGGCGCGCTCGCCACCGAA 78756

Qy        706 GACCAACTCCTCACCCCCCACTACTGGACCCAACAAGCCCGCAACACCGTCGACTACGCC 765
              ||| |||| | | | ||| ||||||||    | || | |||| | ||||||  || |
Db      78755 GACGAACTGCGCTCGCCCGACTACTGGGTGCGGCAGGTCCGCGAAACCGTCCGCTTCTGT 78696

Qy        766 ACCACCACCCAAACCCTCCACCAACACGGCGTCACCACCTACATCGAACTCGGACCCGAC 825
```

```
Db      78695 GCTGCCGTGCGCACCCTTGAGGCCGAGGGCGTCACCACCTTCGTGGAGATCGGCACCGGC 78636

Qy        826 AACACCCTCACCACCCTCACCCACCACAACCTCCCCAACCCCCCCACCACCACCCTCACC 885

Db      78635 GGCGTCCTCACCCCCATGGTCCAGGACTGTCTGACCACCCTCGAAGAGCCCGTTCTCGTC 78576

Qy        886 CTCACCCACCCCCACCACCACCCCCAAACCCACCTCCTC 924

Db      78575 CCTCTGCTGCGCACCGGCCGCCCCGAAACCGTCGCTCTC 78537


RESULT 14
AX211705
LOCUS       AX211705               65140 bp    DNA     linear   PAT 06-SEP-2001
DEFINITION  Sequence 1 from Patent WO0159126.
ACCESSION   AX211705
VERSION     AX211705.1  GI:15523937
KEYWORDS    .
SOURCE      Streptomyces noursei.
  ORGANISM  Streptomyces noursei
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1  (bases 1 to 65140)
  AUTHORS   Zotchev,S.B., Sekurova,O.N., Fjaervik,E., Brautaset,T.,
            Stroem,A.R., Valla,S., Ellingsen,T.E., Sletta,H.V. and
            Gulliksen,O.M.
  TITLE     Gene cluster encoding a nystatin polyketide synthase and its
            manipulation and utility
  JOURNAL   Patent: WO 0159126-A 1 16-AUG-2001;
            Norges Teknisk Naturvitenskapelige Universitet (NO) ; STIFTELSEN
            IND OG TEKNISK FORSKNING VED NORGES TEKNISKE HOGSKOLE (NO) ;
            ALPHARMA AS (NO) ; Sinvent AS (NO) ; Zotchev, Sergey Borisovich
            (NO) ; Sekurova, Olga Nikolayivna (NO) ; Fjaervik, Espen (NO) ;
            Brautaset, Trygve (NO) ; Strom, Arne Reidar (NO) ; Valla, Svein
            (NO)
FEATURES             Location/Qualifiers
     source          1. .65140
                     /organism="Streptomyces noursei"
                     /db_xref="taxon:1971"
                     /note="ATCC 11455"
BASE COUNT     8270 a  25171 c  22273 g  9426 t
ORIGIN

  Query Match             32.5%;  Score 300.6;  DB 6;  Length 65140;
  Best Local Similarity   62.1%;  Pred. No. 1.3e-23;
  Matches  581;  Conservative    0;  Mismatches  309;  Indels   45;  Gaps    5;

Qy          1 ACCGCATTCATCTGCTCCGGACAGGGCACCCAACGCCCCGGCATGGCCCACGGCCTCTAC 60

Db      23408 ACCGCCTTCCTCTTCTCCGGCCAGGGCTCCCAACGGCTCGGCATGGGCCGCGTCCTCTAC 23467

Qy         61 CACACCCACCCCGTCTTCGCCGCCGCACTCAACGACATCTGCACCCACCTCGACCCCCAC 120

Db      23468 GAGCGGTTCCCCGCCTTCGCCGAAGCCCTCGACACCGTCCTCACCGCCCTCGACGCGGAA 23527

Qy        121 CTCGACCACCCCCTCCTCCCCCTCCTCACCCAAAACGACAACGACAACGAGGACGCGGCC 180

Db      23528 CTCGGCCACCCCCTCC----------------GCGACATCATCTGGGGCGAGGACGCT 23569

Qy        181 GCACTGCTCCAGCAGACCCGCTACGCCCAGCCCGCCCTCTTCGCCTTCCAGGTCGCCCTC 240

Db      23570 CAACTCGTCGACCGGACCGGCTACACCCAACCCGCCCTGTTCGCCATCGAGGTGGCACTC 23629

Qy        241 CACCGCCTCCTCACCGACGGCTACCACATCACCCCCCACTACTACGCCGGACACTCCCTC 300

Db      23630 TTCCGCCTCCT---GGAAGCCTGGGGCATCACACCGGACTTCGTGGCCGGCCACTCCATC 23686

Qy        301 GGCGAAATCACCGCCGCCCACCTCGCCGGCATCCTCACCCTCACCGACGCCACCACCCTC 360

Db      23687 GGCGAGATCGCCGCCGCACACGTCGCCGGCGTGCTCTCCCTCGGCGACGCCTGCCGCCTC 23746

Qy        361 ATCACCCAACGCGCCACCCTCATGCAAACCATGCCC---CCCGGCACCATGACCACCCTC 417

Db      23747 GTCGTGGCCCGCGCCGTGCTGATGCAGTCGCTGCCCGAAGGCGGCGCGATGATCGCCGTC 23806

Qy        418 CACACCACCCCCCACCACATCACCCACCACCTCACCGCCCACGAAAACGACCTCGCCATC 477

Db      23807 CAGGCCACCGAGGACGAGGTCCTGCCCCTCCTCACCG---------ACGACGTCTCGATC 23857

Qy        478 GCCGCCATCAACACCCCCACCTCCCTCGTCATCAGCGGCACCCCCCACACCGTCCAACAC 537

Db      23858 GCCGCCGTCAACAGCCCGACCTCCGTCGTCGTCTCCGGCTACGAGAACGCCACCCTCGCC 23917

Qy        538 ATCACCACCCTCTGCCAACAACAAGGCATCAAAACCAAAACCCTCCCCACCAACCACGCC 597

Db      23918 GTCGCCCGGCACTTCGCCGACCAGGGCCGCCGCACCACGCGGCTGCGCGTCAGCCACGCC 23977

Qy        598 TTCCACTCCCCCCACACCAACCCCATCCTCAACCAACTCCACCAGCACACCCAAACCCTC 657

Db      23978 TTCCACTCGCCGCTGATGGCGCCGATGCTCGACGACTTCCGCGCCGTCGTCGAGAGCCTC 24037

Qy        658 ACCTACCACCCACCCCACACCCCCCTCATCACCGCCAACA----------CCCCACCC 705

Db      24038 ACCTTCACCGCCCCCACGACCCCCGTCGTCTCCAACCTGACCGGCGAACTGGCCCCGGCC 24097

Qy        706 GACCAACTCCTCACCCCCCACTACTGGACCCAACAAGCCCGCAACACCGTCGACTACGCC 765

Db      24098 GAGGCGCTCTGCTCGGCCGACTACTGGGTCCGGCACGTCCGCGAGGCGGTCCGCTTCGCC 24157

Qy        766 ACCACCACCCAAACCCTCCACCAACACGGCGTCACCACCTACATCGAACTCGGACCCGAC 825

Db      24158 GACGGCATCCGCACCCTCGCCGACCGCGGCGTCACCACCTTCGTCGAACTCGGCCCCGAC 24217

Qy        826 AACACCCTCACCACCCTCACCCACCACAACCTCCCCAACCCCCCCACCACCACCCTCACC 885

Db      24218 AGCGTGCTGTCCGCCATGGCCCAGGAGTCCGCCCCCGAAGGCGCCGGCACCATCCCGCTC 24277

Qy        886 CTCACCCACCCCCACCACCACCCCCAAACCCACCT 920

Db      24278 CTGCGCCGCGACCGGCCCGAGGAACAGGCCGTCCT 24312


RESULT 15
AF263912
LOCUS       AF263912               123580 bp    DNA     linear   BCT 24-MAY-2000
DEFINITION  Streptomyces noursei ATCC 11455 nystatin biosynthetic gene cluster,
            complete sequence.
ACCESSION   AF263912
VERSION     AF263912.1  GI:8050835
KEYWORDS    .
SOURCE      Streptomyces noursei.
  ORGANISM  Streptomyces noursei
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1  (bases 1 to 123580)
  AUTHORS   Brautaset,T., Sekurova,O.N., Sletta,H., Ellingsen,T.E., StrLm,A.R.,
            Valla,S. and Zotchev,S.B.
  TITLE     Biosynthesis of the polyene antifungal antibiotic nystatin in
            Streptomyces noursei ATCC 11455: analysis of the gene cluster and
            deduction of the biosynthetic pathway
  JOURNAL   Chem. Biol. 7 (6), 395-403 (2000)
  MEDLINE   20334850
   PUBMED   10873841
REFERENCE   2  (bases 1 to 123580)
  AUTHORS   Brautaset,T., Sekurova,O.N., Sletta,H., Ellingsen,T.E., Strom,A.R.,
            Valla,S. and Zotchev,S.B.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-MAY-2000) Unigen, NTNU, O. Kyrres gt. 3, Trondheim
            N-7489, Norway
FEATURES             Location/Qualifiers
     source          1. .123580
                     /organism="Streptomyces noursei"
                     /strain="ATCC 11455"
                     /db_xref="ATCC:11455"
                     /db_xref="taxon:1971"
```

```
gene            complement(46. .783)
                /gene="nysF"
CDS             complement(46. .783)
                /gene="nysF"
                /function="presumably post-translationally modifies ACP
                domains on PKS"
                /note="putative 4'-phosphopantheteine transferase"
                /codon_start=1
                /transl_table=11
                /product="NysF"
                /protein_id="AAF71762.1"
                /db_xref="GI:8050836"
                /translation="MIELILPATVATEAAYDDRPRPGDRLLSSEREVIARAVESRQRE
                FTTVRHLARRALRRLGHPDRAILPNRRGAPQWPPGIVGSMTHCAGYRAAAVSPAELSA
                AVSIDAEPNGPLPAGVLNAIALPSERPHLVALAAHRPDVHWDRLLFSAKESVFKAWYP
                LTQRELDFSEAEIVIDPTQGAFTARLLVPGPLLGGRRVTVFPGRWHSTPALLTTAVHL
                PAPTPRRDREHRTHLTVNSPLPRPTFG"
gene            complement(867. .2684)
                /gene="nysG"
CDS             complement(867. .2684)
                /gene="nysG"
                /function="presumably involved in efflux of nystatin"
                /note="putative transporter (ABC family)"
                /codon_start=1
                /transl_table=11
                /product="NysG"
                /protein_id="AAF71763.1"
                /db_xref="GI:8050837"
                /translation="MASPDDLEEERTAPRPARRLVGLLRPHRRSVALAVSMGVGGIVL
                NAFGPLLLGRVTDLIADGVLGGVPGPAPGIDFAAIGRLLLVLLALYVVASLFMLAQGR
                LVASAVWRTIHELRRDAREKLTRLPLRHFDRQPAGELLSRTTNDIDNLQQTLQQTLAE
                LITSIFSLLTMLVLMLVISPSLAVVMLLSVPVSALIAARISKRAQPHYAAQWSANGTL
                NAHVEEVCTGHALIKGFDRRAAAEERFDACNDAVYRAAAKAQFASGAMEPVMMFVANL
                GYVLVAVIGAWKVINGTLTLGDVQAFILYARQFSQPIVEIASVAGRLQSGIASAQRVF
                TLLDAPEQAPDPLRPGTPARAEGRVEFTDVSFRYSPDTPLIENLSLTVEPGSTVAIVG
                PTGAGKTTLGNLLMRFYEPDSGRILLDGTDTATMTRDDLRSRFGLVLQDTWLFGGTIA
                ENIAYGAPGACRADIEEAARATCADRFIRTLPQGYDTVLDDESGTVSAGEKQLLTVAR
                AFLARPAVLVLDEATSSVDTRTEVLIQRAMNSLRAGRTSFVIAHRLSTIRDADLIVVM
                DAGRIVEQGTHDQLLCAQGLYARLHAARTHTPTAGAAAG"
gene            complement(2662. .4416)
                /gene="nysH"
CDS             complement(2662. .4416)
                /gene="nysH"
                /function="presumably involved in efflux of nystatin"
                /note="putative transporter (ABC family)"
                /codon_start=1
                /transl_table=11
                /product="NysH"
                /protein_id="AAF71764.1"
                /db_xref="GI:8050838"
                /translation="MLLRLLRAQLRPYAWATAALVALQLVQILGTLLLPTLGAALIDQ
                GVVRGDGGRITELGVVMGVVALVQIAAALGAAALAARTATAMGRDLRSALFRRILDFS
                AREIGRFGTPSLLTRSVNDVQQVQNLAQTGFGIVVCAPLMCLGSVLLALRQDVPLALL
                LVALVLVVAVCFGLLLARMGTLYARMQLTLDRLGRLLREAITGVRVVRSFVRDDHERA
                RFAQTNDAFLVVSRRVGRLIATMLPVVLLLMNGFTVALLWTGSHRIDAGRMPIGSLSA
                LLSYLSLILMSVVMLAFVFLSVPRARVCAGRIAEVLDTGSSVAPPAAPQPVRGPAGRI
                ELCAAGYRYPGAEEPVLRDVDLTVEPGERIAVLGSTGSGKTTLLNLVLRLADATEGAV
                RVGGTDVRELTAATLAAAVGFVPQRPYLFSGTVASNLRFGRPDATDEELWEALRVAQA
                ADFVARMPDGLDAEITQGGGNVSGGQRQRLSLARALLRRPEIYLFDDCFSALDQATDA
                ALRTALVPYTAGATVITVAQRISAGRDADRIVVLDRGRVVAQGTHDVLLRTSPTYREI
                ALSQLTEEEAAHGLAGRP"
gene            4714. .5748
                /gene="nysDIII"
CDS             4714. .5748
                /gene="nysDIII"
                /function="presumably involved in mycosamine biosynthesis"
                /note="putative dGDP-mannose-4,6-dehydratase"
                /codon_start=1
                /transl_table=11
                /product="NysDIII"
                /protein_id="AAF71765.1"
                /db_xref="GI:8050839"
                /translation="MSKRALITGITGQDGSYLAEHLLSQGYQVWGLIRGQANPRKSRV
                SRLASELDFIDGDLMDQGSLVSAVDTVQPDEVYNLGAISFVPMSWQQAELVTEVNGMG
                VLRMLEAIRMVSGLSTSRTVSPRGQIRFYQASSSEMFGKAAETPQRETTLFHPRSPYG
                AAKAYGHYITRNYRESFGMYAVSGMLFNHESPRRGQEFVTRKISLAVARIKQGLQDKL
                ALGNLDAVRDWGYAGDYVRAMHLMLQQDAGDDYVIGTGQMHSVRDAVRIAFEHVGLNW
                EDYVVIDPDLVRPAEVEVLCADSAKAQDRLGWKPDVDFPTLMRMMVDSDLAQVSRENQ
                YGDVLLAANW"
gene            5930. .34363
                /gene="nysI"
CDS             5930. .34363
                /gene="nysI"
                /function="responsible for condensation steps 9 to 14 in
                the nystatin polyketide backbone synthesis"
                /note="polyketide synthase"
                /codon_start=1
                /transl_table=11
                /product="NysI"
                /protein_id="AAF71766.1"
                /db_xref="GI:8050840"
                /translation="MDNEQKLRDYLKLATADLRRTRRRVHKLESAAQEPVAIIGMTCR
                YPGGVRSPEDLWRMVEAGEHGVTPFPTDRGWDLEALAAAPTASGGFLHDAPDFDADFF
                GISPREAVAMDPQQRVVLESAWEAFERAGIDPTSVKGSRTGVFIGAMAQDYRVGPADG
                AEGFQLTGNTGSVLSGRISYTFGTVGPAVTVDTACSSSLVAVHLATQALRAGECTLAL
                AGGVTIMSGPGTFIEMGRQGGLSADGRCRSFGDTADGTGWAEGVGILVLERLSDAVRN
                GHEILAVVRGTAVNQDGASNGLTAPNGPSQQQVIQQALVNARLAAGDIDVVEAHGTGT
                TLGDPVEAQALLATYGQNRPADRPLLLGSVKSNLSHTQAAAGVAGVIKMVMAMRHGTL
                PRTLHAEEPTHHVDWSQGAVRLLTDTTDWPATGAPRRAAVSSFGISGTNAHTIIEQAP
                EPQPEDAATAQDDAAGSTPATAPVVPGVVPVLLSGRTPDALRGQAAALRAALDTGRRP
                DLLDLAHSLATTRAGFEHRAVLLATDHPALTDGLTALADADDPAAAPAWITGTTRAET
                RLAVLFTGQGAQRLGAGRELAARFPAFATALDAALDAFTPHLDRPLREVLWGTDAALL
                DRTAYAQPALFAVEVALYRLIESFGVRPDHLAGHSVGEIVAAHLAGVLSLADAATLVA
                ARGRLMQALPDGGAMIAVQASEADVAPLLAGHEDQVAIAAVNGPSAVVLSGAEATVTA
                LAEQLAADGRKTRRLRVSHAFHSPLMEPMLDAFRAVVEDLTLQPPLLPVVSNLTGKPA
                TVAQLTSADYWVDHVRHAVRFADGIDWLARHDTTAFLELGPDGVLSAMAQDCLDAADA
                DAVTLPALRAGRPEEHTLTTALAGLHVHGATLDWTGCFAGTGARRTDLPTYAFQRRRY
                WPKALQSGTADLRSVGLGAAHHPLLSAAVSLADAGGTLLTGRLSRQTHPWLADHTVRG
                TTLLPGTAFLELAVRAGDEVGCDRVEELTLAAPLLLPEQGGVQVQLWIGNPDVSGRRT
                VNVHARPDTGDDTPWTAHATGVLTTADASRQLPASSEQGGTPLAGDPHPALDAAQWPP
                AGAEPLPLDGHYDRLADGGFGYGPVFQGLRAAWRGGDVVYAEVELPEAGRSDAEAFGL
                HPALLDAALHAAPFTGLGERGRGGLPFSWEGVSLHAGGATTLRVRLTPVADDALALTV
                ADGTGAPVLSVDSLVLRSVATQQLDTAAAVARDALFRLDWTPVQPTATDPGPVALLGA
                DPFGLLTHAGFADAPAYPDLAALAAADGPVPTTVVLSLAGTGDDAADPARSAHRCAAE
                ALAAVQTWLDHHERFAAARLVFVTRGATVGRDVAAAAVWGLVRSAQSENPGCFALVDL
                DPDGAVGAAALVAALVSGEPQLAVRGDVLRVARLVRRPLTEVGAGADGTGDGVGDGSG
                VSFSGEGAVLVTGGTGGLGAVLARHLVAEYGVRDLLLVSRSGERAVGAGELVAELAGV
                GARVRVVACDVTDRAAVVELVGGHAVSAVVHAAGVLDDGMVGALTGERLSAVLRPKVD
                AVWHLHEATRGLDLDAFVVFSSLAGVFGSPGQANYAAANAFLDALMTRRRAEGLPGLS
                LAWGPWEQSGGMTGTLTDVDAERLARSGVPPLSVAQGLALFDAAVAGTDATCVPVRLD
                LPVLRARGEVPPLLRSLIRVRARRAAVAGSATAGNLAQRLRRLDEDGRDEMVLDLVRG
                QVALVLGHATGGDVDAGRAFRDLGFDSLTAVELRNRLNTVTGLRLPATLVFDYPTVRH
                LATYVLDELLGTDAEVATVQPAAAVAVADDPIVIVGMACRYPGGVSSPEDLWRVLTEGT
                DAVSGFPTNRGWDVESLYHPDPDHPGTSYTRSGGFLHEAGEFDPGFFGMSPREALATD
                SQQRLLLESSWEAIERAGIDPVSLRGSRTGVFAGVMYSDYSAMLASPEFEGFQGSGSS
                PSLASGRVAYTLGLEGPAVTVDTACSSSLVAMHWAMQALRSGECGLALAGGVTVMSTP
                AVFVDFARQRGLSPDGRCKAFADAADGVGWSEGVGVLVLERQSDAVRNGHEILAVVRG
                SAVNQDGASNGLTAPNGPSQQRVIRQALASGGLTAGDVDVVEAHGTGTTLGDPIEAQA
                LLATYGRDREPERPLLLGSVKSNLGHTQAAAGVAGVIKMVLAMRHGVVPRTLHVDAPS
                SHVDWSEGAVELLSEQAAWPETGRVRRAGVSSFGISGTNVHVIVEQAPGAKAIAAAGA
                ARRTPGAVPVLLSGRGRSALRGQAARLLGHLQARPDAELVDVALSLATTRSRFEQRAA
                VVAQDRDQLIASLGALAADRPDPAVVEGEAAGRGRTAVLFTGQGSQRAAMGRELHEVQ
                PEFAAAFDAVCAVFDPLLDRPLREVVFAEDGSDEAALLDETGWTQPALFAVEVALFRL
                VESWGVRPDFVAGHSIGEIAAAHVAGVLTLEDACRLVAARATLMQALPTGGAMIAIQA
                TEDEIAAHLDDTVAIAAVNGPQSVVISGDEEAAETIAATFAERGRKTKRLRVSHAFHS
                PRMDGMLDAFRIVAEGLTYRAPRIPLVSDLTGRRADDAEVCTAEYWVRHVREAVRFAD
                CVRTLRDAGATTFLELGSDGLLTAMAEDTLGDDHDAELVPMLRAGRAEELAAATALAR
                LQVRGVDVDWAAYLAGTGARRTDLPTYAFQHAYYWPQLPTPAAALAAADPADQQLWAA
                VERGDARELADILGLGEQDLTPLDSLLPALTSWRRGNQEKHLLDTLRYRVEWTRLSKP
                TAPVLDGTWLLVASDATAADQPALLDGLADALGSHGARVRRLLLDDSCADRAVLAERL
                ARTADVDAATQVLSVLPLDERDADDCPPLTRGLALTVALVQALADTGAQGRLWTATRG
                AVSTNPADPVTHPVQAAAWGLGRGVALEHPRLWGGLVDLPQVFDERAGQRLAGILAVK
                DAPDGEDQVALRATGVSGRRLVRHTVEALPTAAEFTATGTVLITGGTGGLGAEVARWL
                ARAGAQHLVLTSRRGPDAPGAAELRAELEGYGPSVSVVACDVADRDALAAVLTALPEE
                LPLTGVVHTAGVGHYGPLDTLSTAEFAGLTAAKLAGAAHLDALLADRELDFFVLFGSI
                AGVWGSGNQSAYGAANAYLDALALHRRARGLAATSVAWGPWAEAGMAADDAVSETLRR
```

```
                     QGLGLLDPAPAMTELRRAVVRQDVTVTVADVDWQRYAPLFTSARPSALIAGLPEVRAL
                     AADERTEQDATGASEVVTRVRALAEPEQLRLLTDLVRTESATVLGHSSADAVPEGRAF
                     RDVGFDSLTAVELRKRLGAATGLSLPSTMVFDYPTPLELAQYLRAEILGAVLEVAGPV
                     ATGGADDEPIAIIGMACRFPGGVSSPEQLWDLVASGTDAISEFPVNRGWQTGHLFDPD
                     PDRPGTTYSTQGGFLHEADEFDPTFFGISPREALVMDPQQRLLLETTWESFERAGIRP
                     ETLRSTLTGTFVGSSYQEYGLGAGDGTEGHMVTGSSPSVLSGRLSYVFGLEGPAVTVD
                     TACSSSLVALHLACQSLRNGESNLAVAGGATIMTTPNPFIAFSRQRALAKDGRCKAFS
                     DDADGMTLAEGVGVVLVERLSDAQRNGHPVLAVLRGSAINQDGASNGLTAPNGPSQQR
                     VIRQALANARLAPGDIDALEAHGTGTPLGDPIEAQALFATYGRDRDPESALLLGSVKS
                     NIGHTQSAAGIASVIKMVMALRHSELPPTLHADAPSSHVDWSAGTVRLLTQARAWPET
                     GRPRRAAVSSFGISGTNAHVLLEQAPVADTPAEERPAVAPVPIAAGVVPWVVTARSAA
                     ALRGQAERLLAHAETVGTALPAAGPLDIGLSLVSARARFEHRAVVVPPAGTDPLAALR
                     AVATDGPSPVVARGVADVEGRTVFVFPGQGSQWVGMGSQLLDESAVFAERIAECAAAL
                     AEFTDWSLVDVLRGVVGAPSLERVDVVQPASFAVMVSLAALWRSRGVLPDAVVGHSQG
                     EIAAAVVSGALSLRDGARVVALRSQAIGRALAGRGGMMSVALSVDVLEPRLVEFEGRV
                     SVAAVNGPRSVVVAGEPEALDALHARLTADDIRARRIAVDYASHSHQVEDLHEELLEV
                     LAELAPRTSEVPFFSTVTGDWLDTARMDAGYWFRNLRGRVRFADAVADLLAAEYRAFV
                     EVSSHPVLSMAVQEAIDEAGVPAVAAGTLRRDQGGTDRFLLSAAEVFVRGVDVDWAGL
                     FEGTGASRIDLPTYAFQHEHLWAVPPAPEAVAAADPDDAAFWTAVEDGDVSALTAALG
                     TDEDSVAAVLPALTSWRRARRDRSTVDAWRYRVAWKPLGGTLPHPSLTGTWLLVTADG
                     IDDTDVAGALETYGAEVRRLVLDEECVDRAVLRERLAGAEDVTGIVSVLAAAERTDAV
                     PGTSLVLGTALTVALIQALGDAEIDAPVWALTRGAVSTGRADELTAPVQAQVTGIGWT
                     AALEHPQRWGGTLDLPAALDARAAQRLAAVLSGALGSDDQLAIRPSGVFTRRIVRAEA
```

Query Match 32.5%; Score 300.6; DB 1; Length 123580;
Best Local Similarity 62.1%; Pred. No. 1.1e-23;
Matches 581; Conservative 0; Mismatches 309; Indels 45; Gaps 5;

```
Qy          1 ACCGCATTCATCTGCTCCGGACAGGGCACCCAACGCCCCGGCATGGCCCACGGCCTCTAC 60
              ||||| ||| ||| |||||| |||||| ||||||| | |||||||| || || |||||||
Db      83261 ACCGCCTTCCTCTTCTCCGGCCAGGGCTCCCAACGGCTCGGCATGGGCCGCGTCCTCTAC 83320

Qy         61 CACACCCACCCCGTCTTCGCCGCCGCACTCAACGACATCTGCACCCACCTCGACCCCCAC 120
                |     ||||| ||||||||  || ||| ||  | ||  ||||  ||||||| |   |
Db      83321 GAGCGGTTCCCCGCCTTCGCCGAAGCCCTCGACACCGTCCTCACCGCCCTCGACGCGGAA 83380

Qy        121 CTCGACCACCCCCTCCTCCCCCTCCTCACCCAAAACGACAACGACAACGAGGACGCGGCC 180
              |||| ||||||||||                |||||  |   |   |  || |  ||
Db      83381 CTCGGCCACCCCCTCC------------------GCGACATCATCTGGGGCGAGGACGCT 83422

Qy        181 GCACTGCTCCAGCAGACCCGCTACGCCCAGCCCGCCCTCTTCGCCTTCCAGGTCGCCCTC 240
                |||  || | | |||| ||||| |||| |||||||| |||||| || |||| || |||
Db      83423 CAACTCGTCGACCGGACCGGCTACACCCAACCCGCCCTGTTCGCCATCGAGGTGGCACTC 83482

Qy        241 CACCGCCTCCTCACCGACGGCTACCACATCACCCCCCACTACTACGCCGGACACTCCCTC 300
              |||||||||    || | ||    |||||| ||  ||| |   ||||| |||||| ||
Db      83483 TTCCGCCTCCT---GGAAGCCTGGGGCATCACACCGGACTTCGTGGCCGGCCACTCCATC 83539

Qy        301 GGCGAAATCACCGCCGCCCACCTCGCCGGCATCCTCACCCTCACCGACGCCACCACCCTC 360
              ||||| ||| ||||||| ||| |||||||| | ||| |||||  |||||||  |  ||||
Db      83540 GGCGAGATCGCCGCCGCACACGTCGCCGGCGTGCTCTCCCTCGGCGACGCCTGCCGCCTC 83599

Qy        361 ATCACCCAACGCGCCACCCTCATGCAAACCATGCCC---CCCGGCACCATGACCACCCTC 417
                ||      ||||||   || |||||  |. |||||     |||| | |||| | || ||
Db      83600 GTCGTGGCCCGCGCCGTGCTGATGCAGTCGCTGCCCGAAGGCGGCGCGATGATCGCCGTC 83659

Qy        418 CACACCACCCCCCACCACATCACCCACCACCTCACCGCCCACGAAAACGACCTCGCCATC 477
              ||  |||||    || |  ||   | || ||||||||         |||||  || | |||
Db      83660 CAGGCCACCGAGGACGAGGTCCTGCCCCTCCTCACCG---------ACGACGTCTCGATC 83710

Qy        478 GCCGCCATCAACACCCCCACCTCCCTCGTCATCAGCGGCACCCCCCACACCGTCCAACAC 537
              |||||| |||||| ||| |||||| ||||| ||  ||||  |     || ||  ||    |
Db      83711 GCCGCCGTCAACAGCCCGACCTCCGTCGTCGTCTCCGGCTACGAGAACGCCACCCTCGCC 83770

Qy        538 ATCACCACCCTCTGCCAACAACAAGGCATCAAAACCAAAACCCTCCCCACCAACCACGCC 597
              || ||   | || |   | || ||| |   ||||     || | |  || |||||||
Db      83771 GTCGCCCGGCACTTCGCCGACCAGGGCCGCCGCACCACGCGGCTGCGCGTCAGCCACGCC 83830

Qy        598 TTCCACTCCCCCCACACCAACCCCATCCTCAACCAACTCCACCAGCACACCCAAACCCTC 657
              |||||||| || |  |    || || ||| || |  ||| |     |  | | | ||||
Db      83831 TTCCACTCGCCGCTGATGGCGCCGATGCTCGACGACTTCCGCGCCGTCGTCGAGAGCCTC 83890

Qy        658 ACCTACCACCCACCCCACACCCCCCTCATCACCGCCAACA------------CCCCACCC 705
```

```
              |||| |  | | |||   |||||| || || ||  |   |            ||||  ||
Db      83891 ACCTTCACCGCCCCCACGACCCCCGTCGTCTCCAACCTGACCGGCGAACTGGCCCCGGCC 83950

Qy        706 GACCAACTCCTCACCCCCCACTACTGGACCCAACAAGCCCGCAACACCGTCGACTACGCC 765
              ||     |||  | |  || |||||||  ||  || | |||| |  | |||  || ||||
Db      83951 GAGGCGCTCTGCTCGGCCGACTACTGGGTCCGGCACGTCCGCGAGGCGGTCCGCTTCGCC 84010

Qy        766 ACCACCACCCAAACCCTCCACCAACACGGCGTCACCACCTACATCGAACTCGGACCCGAC 825
                |  || ||  ||||||  | | | |||||||||||||| | ||||||||| ||||||
Db      84011 GACGGCATCCGCACCCTCGCCGACCGCGGCGTCACCACCTTCGTCGAACTCGGCCCCGAC 84070

Qy        826 AACACCCTCACCACCCTCACCCACCACAACCTCCCCAACCCCCCCACCACCACCCTCACC 885
              | |   ||  || || |  ||||  |   |  |||| |   | ||  |||| ||    |
Db      84071 AGCGTGCTGTCCGCCATGGCCCAGGAGTCCGCCCCCGAAGGCGCCGGCACCATCCCGCTC 84130

Qy        886 CTCACCCACCCCCACCACCACCCCCAAACCCACCT 920
              ||   || |  ||  | | |    ||  ||  |||
Db      84131 CTGCGCCGCGACCGGCCCGAGGAACAGGCCGTCCT 84165
```

Search completed: June 17, 2003, 17:55:47
Job time : 1798.83 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 17, 2003, 11:26:45 ; Search time 183.808 Seconds
(without alignments)
11320.736 Million cell updates/sec

Title: US-09-914-286-1_COPY_7906_8829
Perfect score: 924
Sequence: 1 accgcattcatctgctccgg.....acccccaaacccacctcctc 924

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_101002:*
1: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	924	100.0	11916	22	AAH79279	Streptomyces averm
2	924	100.0	30690	21	AAA92301	S. avermitilis ave
3	924	100.0	30690	22	AAH79277	Streptomyces averm
4	919.2	99.5	12381	21	AAZ58381	Streptomyces averm
5	873.4	94.5	31422	21	AAA92302	S. avermitilis ave
6	873.4	94.5	31422	22	AAH79278	Streptomyces averm
7	302.2	32.7	20394	22	AAF24892	Pimaricin biosynth
8	300.6	32.5	65140	22	AAD17184	Streptomyces nours
9	300.6	32.5	125401	22	AAD17186	Streptomyces nours
10	273.4	29.6	27541	22	AAD17185	Streptomyces nours
11	266.2	28.8	28598	17	AAT06769	Sorangium cellulos
12	266.2	28.8	28958	18	AAT89956	Sorangium cellulos
13	266.2	28.8	28958	21	AAA75299	DNA sequence of So
14	266.2	28.8	49377	19	AAV05287	The soraphen biosy
15	253	27.4	925	20	AAV99229	DNA encoding an ac
16	253	27.4	925	21	AAA72501	LigAT2 domain from
17	239.4	25.9	15872	18	AAT68715	Streptomyces venez
18	239.4	25.9	15872	21	AAZ87283	S. venezuelae vep
19	238	25.8	3978	21	AAC55785	Type I polyketide
20	238	25.8	12249	21	AAC55840	Complete Mitomycin
21	238	25.8	18331	21	AAC55857	Complete nucleotid
22	232.6	25.2	50937	21	AAA09469	Streptococcus olea
23	220.2	23.8	13842	21	AAZ87297	S. venezuelae macr
24	220.2	23.8	36778	21	AAZ87318	S. venezuelae pik
25	220.2	23.8	37948	21	AAZ87285	S. venezuelae pik
26	220.2	23.8	38506	21	AAA75633	Nucleotide sequenc
27	220.2	23.8	38506	21	AAZ56001	Recombinant cosmid
28	219.4	23.7	43280	18	AAT80413	Tylactone synthase
29	218.8	23.7	1030	20	AAV99230	DNA encoding an ac
30	218.8	23.7	1030	21	AAA72514	Nucleotide sequenc
c 31	211.2	22.9	1075	22	ABA48937	Human breast cell
c 32	211.2	22.9	1075	22	ABA66852	Human foetal liver
c 33	211.2	22.9	1075	22	ABA33921	Probe #12387 for g
c 34	211.2	22.9	1075	22	AAK15288	Human brain expres
c 35	211.2	22.9	1075	22	AAK41011	Human bone marrow
c 36	211.2	22.9	1075	22	AAI21781	Probe #11714 for g
c 37	211.2	22.9	1075	22	AAI47064	Probe #15750 used
c 38	211.2	22.9	1075	22	AAI07466	Probe #7457 used t
c 39	211.2	22.9	1075	24	ABS14983	Human genome-deriv
c 40	211.2	22.9	1403	22	ABA43834	Human breast cell
c 41	211.2	22.9	1403	22	ABA54290	Human foetal liver
c 42	211.2	22.9	1403	22	ABA24047	Probe #2513 for ge
c 43	211.2	22.9	1403	22	AAK02574	Human brain expres
c 44	211.2	22.9	1403	22	AAK28010	Human bone marrow
c 45	211.2	22.9	1403	22	AAI12593	Probe #2526 for ge

ALIGNMENTS

RESULT 1
AAH79279
ID AAH79279 standard; DNA; 11916 BP.
XX
AC AAH79279;
XX
DT 04-DEC-2001 (first entry)
XX
DE Streptomyces avermitilis coding sequence derivative SEQ ID NO: 3.
XX
KW Avermectin aglycone synthase; AAS; avermectin derivative;
KW drug production; vetinary drug; pesticide; ds.
XX
OS Synthetic.
XX
FH Key Location/Qualifiers
FT CDS 1..11916
FT /*tag= a
FT /product= "AAG65268"
FT /partial
XX
PN WO200162939-A1.
XX
PD 30-AUG-2001.
XX
PF 23-FEB-2001; 2001WO-JP01381.
XX
PR 24-FEB-2000; 2000JP-0047405.
XX
PA (KYOW) KYOWA HAKKO KOGYO KK.
PA (KITA) KITASATO INST.
XX

```
PI   Endo H,  Yamaguchi H,  Kanda Y,  Hashimoto S,  Omura S,  Ikeda H;
XX
DR   WPI; 2001-582053/65.
DR   P-PSDB; AAG65268.
XX
PT   New modified avermectin aglycone synthase derived from Streptomyces
PT   avermectilis used in production of 22,23-dihydroavermectin B1a used in
PT   drugs and pesticides -
XX
PS   Claim 10; Page 149-167; 257pp; Japanese.
XX
CC   The present invention relates to the production of modified derivatives
CC   of avermectin aglycone synthase (AAS) derived from Streptomyces
CC   avermitilis. The activity of an acyl carrier protein (ACP),
CC   beta-ketoacyl carrier protein synthase (KS), acyltransferase (AT),
CC   beta-ketoacyl carrier protein reductase (KR), dehydratase (DH), enoyl
CC   reductase (ER) and/or thioesterase (TE) domain may be reduced or
CC   suppressed. The process can be used in the production of drugs, vetinary
CC   drugs and pesticides. The present sequence is modified version of a
CC   fragment of the S. avermitilis genome.
XX
SQ   Sequence 11916 BP; 1831 A; 4428 C; 3820 G; 1837 T; 0 other;

  Query Match             100.0%;  Score 924;  DB 22;  Length 11916;
  Best Local Similarity   100.0%;  Pred. No. 4e-127;
  Matches  924;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ACCGCATTCATCTGCTCCGGACAGGGCACCCAACGCCCCGGCATGGCCCACGGCCTCTAC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7906 ACCGCATTCATCTGCTCCGGACAGGGCACCCAACGCCCCGGCATGGCCCACGGCCTCTAC 7965

Qy         61 CACACCCACCCCGTCTTCGCCGCCGCACTCAACGACATCTGCACCCACCTCGACCCCCAC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7966 CACACCCACCCCGTCTTCGCCGCCGCACTCAACGACATCTGCACCCACCTCGACCCCCAC 8025

Qy        121 CTCGACCACCCCCTCCTCCCCCTCCTCACCCAAAACGACAACGACAACGAGGACGCGGCC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8026 CTCGACCACCCCCTCCTCCCCCTCCTCACCCAAAACGACAACGACAACGAGGACGCGGCC 8085

Qy        181 GCACTGCTCCAGCAGACCCGCTACGCCCAGCCCGCCCTCTTCGCCTTCCAGGTCGCCCTC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8086 GCACTGCTCCAGCAGACCCGCTACGCCCAGCCCGCCCTCTTCGCCTTCCAGGTCGCCCTC 8145

Qy        241 CACCGCCTCCTCACCGACGGCTACCACATCACCCCCCACTACTACGCCGGACACTCCCTC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8146 CACCGCCTCCTCACCGACGGCTACCACATCACCCCCCACTACTACGCCGGACACTCCCTC 8205

Qy        301 GGCGAAATCACCGCCGCCCCACCTCGCCGGCATCCTCACCCTCACCGACGCCACCACCCTC 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8206 GGCGAAATCACCGCCGCCCCACCTCGCCGGCATCCTCACCCTCACCGACGCCACCACCCTC 8265

Qy        361 ATCACCCAACGCGCCACCCTCATGCAAACCATGCCCCCCGGCACCATGACCACCCTCCAC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8266 ATCACCCAACGCGCCACCCTCATGCAAACCATGCCCCCCGGCACCATGACCACCCTCCAC 8325

Qy        421 ACCACCCCCCACCACATCACCCACCACCTCACCGCCCACGAAAACGACCTCGCCATCGCC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8326 ACCACCCCCCACCACATCACCCACCACCTCACCGCCCACGAAAACGACCTCGCCATCGCC 8385

Qy        481 GCCATCAACACCCCCACCTCCCTCGTCATCAGCGGCACCCCCCACACCGTCCAACACATC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8386 GCCATCAACACCCCCACCTCCCTCGTCATCAGCGGCACCCCCCACACCGTCCAACACATC 8445

Qy        541 ACCACCCTCTGCCAACAACAAGGCATCAAAACCAAAACCCTCCCCACCAACCACGCCTTC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8446 ACCACCCTCTGCCAACAACAAGGCATCAAAACCAAAACCCTCCCCACCAACCACGCCTTC 8505

Qy        601 CACTCCCCCCACACCAACCCCATCCTCAACCAACTCCACCAGCACACCCAAACCCTCACC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8506 CACTCCCCCCACACCAACCCCATCCTCAACCAACTCCACCAGCACACCCAAACCCTCACC 8565

Qy        661 TACCACCCACCCCACACCCCCCTCATCACCGCCAACACCCCACCCGACCAACTCCTCACC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db       8566 TACCACCCACCCCACACCCCCCTCATCACCGCCAACACCCCACCCGACCAACTCCTCACC 8625

Qy        721 CCCCACTACTGGACCCAACAAGCCCGCAACACCGTCGACTACGCCACCACCACCCAAACC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8626 CCCCACTACTGGACCCAACAAGCCCGCAACACCGTCGACTACGCCACCACCACCCAAACC 8685

Qy        781 CTCCACCAACACGGCGTCACCACCTACATCGAACTCGGACCCGACAACACCCTCACCACC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8686 CTCCACCAACACGGCGTCACCACCTACATCGAACTCGGACCCGACAACACCCTCACCACC 8745

Qy        841 CTCACCCACCACAACCTCCCCAACCCCCCCACCACCACCCTCACCCTCACCCACCCCCAC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8746 CTCACCCACCACAACCTCCCCAACCCCCCCACCACCACCCTCACCCTCACCCACCCCCAC 8805

Qy        901 CACCACCCCCAAACCCACCTCCTC 924
              |||||||||||||||||||||||||
Db       8806 CACCACCCCCAAACCCACCTCCTC 8829


RESULT 2
AAA92301
ID   AAA92301 standard; DNA; 30690 BP.
XX
AC   AAA92301;
XX
DT   10-JAN-2001  (first entry)
XX
DE   S. avermitilis avermectin aglycon synthase DNA aveAI SEQ ID NO:1.
XX
KW   Streptomyces avermitilis; avermectin aglycon synthase; biosynthesis;
KW   multifunctional enzyme; polyketide; avermectin; veterinary drug;
KW   agrochemical; ds.
XX
OS   Streptomyces avermitilis.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..11919
FT                   /*tag=  a
FT                   /note= "avermectin aglycon synthase protein"
FT   CDS             11971..30690
FT                   /*tag=  b
FT                   /note= "avermectin aglycon synthase protein"
XX
PN   WO200050605-A1.
XX
PD   31-AUG-2000.
XX
PF   23-FEB-2000; 2000WO-JP01041.
XX
PR   24-FEB-1999;   99JP-0046961.
XX
PA   (KITA ) KITASATO INST.
XX
PI   Omura S,  Ikeda H;
XX
DR   WPI; 2000-565458/52.
DR   P-PSDB; AAB23749, AAB23750.
XX
PT   Avermectin aglycone synthase DNA and proteins encoded by all or part of
PT   it for the production of avermectin and its derivatives for drug and
PT   agrochemical use  -
XX
PS   Claim 2; Page 66-134; 314pp; Japanese.
XX
CC   The present sequence represents DNA which encodes avermectin aglycon
CC   synthase proteins. Also described are: (1) polypeptides encoded by all
CC   or part of the DNA; (2) expression vectors containing the DNA; (3) host
CC   cells transformed by the vectors; (4) preparation of the polypeptides
CC   by culture of the transformants; (5) preparation of avermectin aglycon
CC   or its derivatives by culture of transformed avermectin-producing
CC   microorganisms; and (6) oligonucleotides of 5-60 bases in length
CC   containing sense or antisense sequences from the avermectin aglycon
CC   synthase DNA. The enzymes are useful for the production of modified
```

```
CC   forms of avermectin and of the intermediates in its biosynthesis, for
CC   use as drugs, veterinary drugs and agrochemicals.
XX
SQ   Sequence 30690 BP; 5356 A; 12454 C; 8617 G; 4263 T; 0 other;

  Query Match           100.0%;  Score 924;  DB 21;  Length 30690;
  Best Local Similarity  100.0%;  Pred. No. 3.7e-127;
  Matches  924;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ACCGCATTCATCTGCTCCGGACAGGGCACCCAACGCCCCGGCATGGCCCACGGCCTCTAC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7906 ACCGCATTCATCTGCTCCGGACAGGGCACCCAACGCCCCGGCATGGCCCACGGCCTCTAC 7965

Qy         61 CACACCCACCCCGTCTTCGCCGCCGCACTCAACGACATCTGCACCCACCTCGACCCCCAC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7966 CACACCCACCCCGTCTTCGCCGCCGCACTCAACGACATCTGCACCCACCTCGACCCCCAC 8025

Qy        121 CTCGACCACCCCCTCCTCCCCCTCCTCACCCAAAACGACAACGACAACGAGGACGCGGCC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8026 CTCGACCACCCCCTCCTCCCCCTCCTCACCCAAAACGACAACGACAACGAGGACGCGGCC 8085

Qy        181 GCACTGCTCCAGCAGACCCGCTACGCCCAGCCCGCCCTCTTCGCCTTCCAGGTCGCCCTC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8086 GCACTGCTCCAGCAGACCCGCTACGCCCAGCCCGCCCTCTTCGCCTTCCAGGTCGCCCTC 8145

Qy        241 CACCGCCTCCTCACCGACGGCTACCACATCACCCCCCACTACTACGCCGGACACTCCCTC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8146 CACCGCCTCCTCACCGACGGCTACCACATCACCCCCCACTACTACGCCGGACACTCCCTC 8205

Qy        301 GGCGAAATCACCGCCGCCCACCTCGCCGGCATCCTCACCCTCACCGACGCCACCACCCTC 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8206 GGCGAAATCACCGCCGCCCACCTCGCCGGCATCCTCACCCTCACCGACGCCACCACCCTC 8265

Qy        361 ATCACCCAACGCGCCACCCTCATGCAAACCATGCCCCCCGGCACCATGACCACCCTCCAC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8266 ATCACCCAACGCGCCACCCTCATGCAAACCATGCCCCCCGGCACCATGACCACCCTCCAC 8325

Qy        421 ACCACCCCCCCACCACATCACCCACCACCTCACCGCCCACGAAAACGACCTCGCCATCGCC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8326 ACCACCCCCCCACCACATCACCCACCACCTCACCGCCCACGAAAACGACCTCGCCATCGCC 8385

Qy        481 GCCATCAACACCCCCACCTCCCTCGTCATCAGCGGCACCCCCCACACCGTCCAACACATC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8386 GCCATCAACACCCCCACCTCCCTCGTCATCAGCGGCACCCCCCACACCGTCCAACACATC 8445

Qy        541 ACCACCCTCTGCCAACAACAAGGCATCAAAACCAAAACCCTCCCCACCAACCACGCCTTC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8446 ACCACCCTCTGCCAACAACAAGGCATCAAAACCAAAACCCTCCCCACCAACCACGCCTTC 8505

Qy        601 CACTCCCCCCACACCAACCCCATCCTCAACCAACTCCACCAGCACACCCAAACCCTCACC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8506 CACTCCCCCCACACCAACCCCATCCTCAACCAACTCCACCAGCACACCCAAACCCTCACC 8565

Qy        661 TACCACCCACCCCACACCCCCCTCATCACCGCCAACACCCCACCCGACCAACTCCTCACC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8566 TACCACCCACCCCACACCCCCCTCATCACCGCCAACACCCCACCCGACCAACTCCTCACC 8625

Qy        721 CCCCACTACTGGACCCAACAAGCCCGCAACACCGTCGACTACGCCACCACCACCCAAACC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8626 CCCCACTACTGGACCCAACAAGCCCGCAACACCGTCGACTACGCCACCACCACCCAAACC 8685

Qy        781 CTCCACCAACACGGCGTCACCACCTACATCGAACTCGGACCCGACAACACCCTCACCACC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8686 CTCCACCAACACGGCGTCACCACCTACATCGAACTCGGACCCGACAACACCCTCACCACC 8745

Qy        841 CTCACCCACCACAACCTCCCCAACCCCCCCACCACCACCCTCACCCTCACCCACCCCCAC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8746 CTCACCCACCACAACCTCCCCAACCCCCCCACCACCACCCTCACCCTCACCCACCCCCAC 8805

Qy        901 CACCACCCCCAAACCCACCTCCTC 924
              |||||||||||||||||||||||||
Db       8806 CACCACCCCCAAACCCACCTCCTC 8829
```

```
RESULT 3
AAH79277
ID   AAH79277 standard; DNA; 30690 BP.
XX
AC   AAH79277;
XX
DT   04-DEC-2001  (first entry)
XX
DE   Streptomyces avermitilis coding sequences SEQ ID NO: 1.
XX
KW   Avermectin aglycone synthase; AAS; avermectin derivative;
KW   drug production; vetinary drug; pesticide; ds.
XX
OS   Streptomyces avermitilis.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..11919
FT                   /*tag=  a
FT                   /product= "AAG65264"
FT                   /partial
FT   CDS             11971..30690
FT                   /*tag=  b
FT                   /product= "AAG65265"
XX
PN   WO200162939-A1.
XX
PD   30-AUG-2001.
XX
PF   23-FEB-2001; 2001WO-JP01381.
XX
PR   24-FEB-2000; 2000JP-0047405.
XX
PA   (KYOW ) KYOWA HAKKO KOGYO KK.
PA   (KITA ) KITASATO INST.
XX
PI   Endo H,  Yamaguchi H,  Kanda Y,  Hashimoto S,  Omura S,  Ikeda H;
XX
DR   WPI; 2001-582053/65.
DR   P-PSDB; AAG65264, AAG65265.
XX
PT   New modified avermectin aglycone synthase derived from Streptomyces
PT   avermectilis used in production of 22,23-dihydroavermectin B1a used in
PT   drugs and pesticides -
XX
PS   Example 2; Page 58-123; 257pp; Japanese.
XX
CC   The present invention relates to the production of modified derivatives
CC   of avermectin aglycone synthase (AAS) derived from Streptomyces
CC   avermitilis. The activity of an acyl carrier protein (ACP),
CC   beta-ketoacyl carrier protein synthase (KS), acyltransferase (AT),
CC   beta-ketoacyl carrier protein reductase (KR), dehydratase (DH), enoyl
CC   reductase (ER) and/or thioesterase (TE) domain may be reduced or
CC   suppressed. The process can be used in the production of drugs, vetinary
CC   drugs and pesticides. The present sequence is a fragment of the S.
CC   avermitilis genome.
XX
SQ   Sequence 30690 BP; 5355 A; 12455 C; 8617 G; 4263 T; 0 other;

  Query Match           100.0%;  Score 924;  DB 22;  Length 30690;
  Best Local Similarity  100.0%;  Pred. No. 3.7e-127;
  Matches  924;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ACCGCATTCATCTGCTCCGGACAGGGCACCCAACGCCCCGGCATGGCCCACGGCCTCTAC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7906 ACCGCATTCATCTGCTCCGGACAGGGCACCCAACGCCCCGGCATGGCCCACGGCCTCTAC 7965

Qy         61 CACACCCACCCCGTCTTCGCCGCCGCACTCAACGACATCTGCACCCACCTCGACCCCCAC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7966 CACACCCACCCCGTCTTCGCCGCCGCACTCAACGACATCTGCACCCACCTCGACCCCCAC 8025

Qy        121 CTCGACCACCCCCTCCTCCCCCTCCTCACCCAAAACGACAACGACAACGAGGACGCGGCC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db       8026 CTCGACCACCCCCTCCTCCCCCTCCTCACCCAAAACGACAACGACAACGAGGACGCGGCC 8085

Qy        181 GCACTGCTCCAGCAGACCCGCTACGCCCAGCCCGCCCTCTTCGCCTTCCAGGTCGCCCTC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8086 GCACTGCTCCAGCAGACCCGCTACGCCCAGCCCGCCCTCTTCGCCTTCCAGGTCGCCCTC 8145

Qy        241 CACCGCCTCCTCACCGACGGCTACCACATCACCCCCCACTACTACGCCGGACACTCCCTC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8146 CACCGCCTCCTCACCGACGGCTACCACATCACCCCCCACTACTACGCCGGACACTCCCTC 8205

Qy        301 GGCGAAATCACCGCCGCCCACCTCGCCGGCATCCTCACCCTCACCGACGCCACCACCCTC 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8206 GGCGAAATCACCGCCGCCCACCTCGCCGGCATCCTCACCCTCACCGACGCCACCACCCTC 8265

Qy        361 ATCACCCAACGCGCCACCCTCATGCAAACCATGCCCCCCGGCACCATGACCACCCTCCAC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8266 ATCACCCAACGCGCCACCCTCATGCAAACCATGCCCCCCGGCACCATGACCACCCTCCAC 8325

Qy        421 ACCACCCCCCACCACATCACCCACCACCTCACCGCCCACGAAAACGACCTCGCCATCGCC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8326 ACCACCCCCCACCACATCACCCACCACCTCACCGCCCACGAAAACGACCTCGCCATCGCC 8385

Qy        481 GCCATCAACACCCCCACCTCCCTCGTCATCAGCGGCACCCCCCACACCGTCCAACACATC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8386 GCCATCAACACCCCCACCTCCCTCGTCATCAGCGGCACCCCCCACACCGTCCAACACATC 8445

Qy        541 ACCACCCTCTGCCAACAACAAGGCATCAAAACCAAAACCCTCCCCACCAACCACGCCTTC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8446 ACCACCCTCTGCCAACAACAAGGCATCAAAACCAAAACCCTCCCCACCAACCACGCCTTC 8505

Qy        601 CACTCCCCCCACACCAACCCCATCCTCAACCAACTCCACCAGCACACCCAAACCCTCACC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8506 CACTCCCCCCACACCAACCCCATCCTCAACCAACTCCACCAGCACACCCAAACCCTCACC 8565

Qy        661 TACCACCCACCCCACACCCCCCTCATCACCGCCAACACCCCACCCGACCAACTCCTCACC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8566 TACCACCCACCCCACACCCCCCTCATCACCGCCAACACCCCACCCGACCAACTCCTCACC 8625

Qy        721 CCCCACTACTGGACCCAACAAGCCCGCAACACCGTCGACTACGCCACCACCACCCAAACC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8626 CCCCACTACTGGACCCAACAAGCCCGCAACACCGTCGACTACGCCACCACCACCCAAACC 8685

Qy        781 CTCCACCAACACGGCGTCACCACCTACATCGAACTCGGACCCGACAACACCCTCACCACC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8686 CTCCACCAACACGGCGTCACCACCTACATCGAACTCGGACCCGACAACACCCTCACCACC 8745

Qy        841 CTCACCCACCACAACCTCCCCAACCCCCCCACCACCACCCTCACCCTCACCCACCCCCAC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8746 CTCACCCACCACAACCTCCCCAACCCCCCCACCACCACCCTCACCCTCACCCACCCCCAC 8805

Qy        901 CACCACCCCCAAACCCACCTCCTC 924
              |||||||||||||||||||||||||
Db       8806 CACCACCCCCAAACCCACCTCCTC 8829


RESULT 4
AAZ58381
ID   AAZ58381 standard; DNA; 12381 BP.
XX
AC   AAZ58381;
XX
DT   23-MAY-2000  (first entry)
XX
DE   Streptomyces avermitilis avermectin polyketide synthase modules 1+2.
XX
KW   Polyketide synthase; avermectin; insecticide; ss.
XX
OS   Streptomyces avermitilis.
XX
PN   WO200001827-A2.
XX
PD   13-JAN-2000.
XX
PF   06-JUL-1999;   99WO-GB02158.
XX
PR   06-JUL-1998;   98GB-0014622.
XX
PA   (BIOT-) BIOTICA TECHNOLOGY LTD.
PA   (PFIZ ) PFIZER INC.
XX
PI   Kellenberger JL,  Leadlay PF,  Staunton J,  Stutzman-Engwall KJ;
PI   McArthur HAI;
XX
DR   WPI; 2000-182117/16.
XX
PT   Mutated Type I polyketide synthase containing a polylinker site in an
PT   extension module for replacement of a reductive loop sequence, for
PT   producing polyketides, e.g. B1 avermectin -
XX
PS   Disclosure; Fig 7a-f; 75pp; English.
XX
CC   The present sequence is that of DNA encoding the first 2 modules
CC   of the avermectin polyketide synthase (PKS) of Streptomyces
CC   avermitilis.  The invention relates to nucleic acids encoding a
CC   Type I PKS such as avermectin in which a polylinker with multiple
CC   restriction sites replaces or 1 more PKS genes encoding enzymes
CC   associated with reduction.  Novel PKS are provided in which in
CC   which the reductive loop in a selected module.of the Type I PKS is
CC   replaced with the equivalent segment from the same or different
CC   PKS gene cluster or by a mutated or synthetic segment.  Vectors and
CC   host cells, and methods for producing novel polyketides by
CC   culturing host cells are claimed.  The polyketides obtained are
CC   useful as antibiotics and insecticides.  Fermentation products
CC   containing C22-C23 dihydroavermectin, ivermectin and B1
CC   avermectins are claimed.
XX
SQ   Sequence 12381 BP; 1884 A; 4561 C; 4005 G; 1931 T; 0 other;

  Query Match             99.5%;  Score 919.2;  DB 21;  Length 12381;
  Best Local Similarity   99.7%;  Pred. No. 2e-126;
  Matches  921;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          1 ACCGCATTCATCTGCTCCGGACAGGGCACCCAACGCCCCGGCATGGCCCACGGCCTCTAC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8723 ACCGCATTCATCTGCTCCGGACAGGGCACCCAACGCCCCGGCATGGCCCACGGCCTCTAC 8782

Qy         61 CACACCCACCCCGTCTTCGCCGCCGCACTCAACGACATCTGCACCCACCTCGACCCCCAC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8783 CACACCCACCCCGTCTTCGCCGCCGCACTCAACGACATCTGCACCCACCTCGACCCCCAC 8842

Qy        121 CTCGACCACCCCCTCCTCCCCCTCCTCACCCAAAACGACAACGACAACGAGGACGCGGCC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8843 CTCGACCACCCCCTCCTCCCCCTCCTCACCCAAAACGACAACGACAACGAGGACGCGGCC 8902

Qy        181 GCACTGCTCCAGCAGACCCGCTACGCCCAGCCCGCCCTCTTCGCCTTCCAGGTCGCCCTC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8903 GCACTGCTCCAGCAGACCCGCTACGCCCAGCCCGCCCTCTTCGCCTTCCAGGTCGCCCTC 8962

Qy        241 CACCGCCTCCTCACCGACGGCTACCACATCACCCCCCACTACTACGCCGGACACTCCCTC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8963 CACCGCCTCCTCACCGACGGCTACCACATCACCCCCCACTACTACGCCGGACACTCCCTC 9022

Qy        301 GGCGAAATCACCGCCGCCCACCTCGCCGGCATCCTCACCCTCACCGACGCCACCACCCTC 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9023 GGCGAAATCACCGCCGCCCACCTCGCCGGCATCCTCACCCTCACCGACGCCACCACCCTC 9082

Qy        361 ATCACCCAACGCGCCACCCTCATGCAAACCATGCCCCCCGGCACCATGACCACCCTCCAC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9083 ATCACCCAACGCGCCACCCTCATGCAAACCATGCCCCCCGGCACCATGACCACCCTCCAC 9142

Qy        421 ACCACCCCCCACCACATCACCCACCACCTCACCGCCCACGAAAACGACCTCGCCATCGCC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9143 ACCACCCCCCACCACATCACCCACCACCTCACCGCCCACGAAAACGACCTCGCCATCGCC 9202

Qy        481 GCCATCAACACCCCCACCTCCCTCGTCATCAGCGGCACCCCCCACACCGTCCAACACATC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db       9203 GCCATCAACACCCCCACCTCCCTCGTCATCAGCGGCACCCCCCACACCGTCCAACACATC 9262

Qy        541 ACCACCCTCTGCCAACAACAAGGCATCAAAACCAAAACCCTCCCCACCAACCACGCCTTC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9263 ACCACCCTCTGCCAACAACAAGGCATCAAAACCAAAACCCTCCCCACCAACCACGCCTTC 9322

Qy        601 CACTCCCCCCACACCAACCCCATCCTCAACCAACTCCACCAGCACACCCAAACCCTCACC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9323 CACTCCCCCCACACCAACCCCATCCTCAACCAACTCCACCAGCACACCCAAACCCTCACC 9382

Qy        661 TACCACCCACCCCACACCCCCCTCATCACCGCCAACACCCCACCCGACCAACTCCTCACC 720
              |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db       9383 TACCACCCACCCCACACCCCCCTCATCACCGACAACACCCCACCCGACCAACTCCTCACC 9442

Qy        721 CCCCACTACTGGACCCAACAAGCCCGCAACACCGTCGACTACGCCACCACCACCCAAACC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9443 CCCCACTACTGGACCCAACAAGCCCGCAACACCGTCGACTACGCCACCACCACCCAAACC 9502

Qy        781 CTCCACCAACACGGCGTCACCACCTACATCGAACTCGGACCCGACAACACCCTCACCACC 840
              |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db       9503 CTCCACCAACACGGCGTCACCACCTACATCGATCTCGGACCCGACAACACCCTCACCACC 9562

Qy        841 CTCACCCACCACAACCTCCCCAACCCCCCCACCACCACCCTCACCCTCACCCACCCCCAC 900
              |||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db       9563 CTCACCCACCACAACCTCCCCAACACCCCCACCACCACCCTCACCCTCACCCACCCCCAC 9622

Qy        901 CACCACCCCCAAACCCACCTCCTC 924
              |||||||||||||||||||||||
Db       9623 CACCACCCCCAAACCCACCTCCTC 9646


RESULT 5
AAA92302
ID   AAA92302 standard; DNA; 31422 BP.
XX
AC   AAA92302;
XX
DT   10-JAN-2001  (first entry)
XX
DE   S. avermitilis avermectin aglycon synthase DNA aveAII SEQ ID NO:2.
XX
KW   Streptomyces avermitilis; avermectin aglycon synthase; biosynthesis;
KW   multifunctional enzyme; polyketide; avermectin; veterinary drug;
KW   agrochemical; ds.
XX
OS   Streptomyces avermitilis.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..14646
FT                   /*tag= a
FT                   /note= "avermectin aglycon synthase protein"
FT   CDS             14824..31422
FT                   /*tag= b
FT                   /note= "avermectin aglycon synthase protein"
XX
PN   WO200050605-A1.
XX
PD   31-AUG-2000.
XX
PF   23-FEB-2000; 2000WO-JP01041.
XX
PR   24-FEB-1999;   99JP-0046961.
XX
PA   (KITA ) KITASATO INST.
XX
PI   Omura S,  Ikeda H;
XX
DR   WPI; 2000-565458/52.
DR   P-PSDB; AAB23751, AAB23752.
XX
PT   Avermectin aglycone synthase DNA and proteins encoded by all or part of
PT   it for the production of avermectin and its derivatives for drug and
PT   agrochemical use  -
XX
PS   Claim 2; Page 134-203; 314pp; Japanese.
XX
CC   The present sequence represents DNA which encodes avermectin aglycon
CC   synthase proteins. Also described are: (1) polypeptides encoded by all
CC   or part of the DNA; (2) expression vectors containing the DNA; (3) host
CC   cells transformed by the vectors; (4) preparation of the polypeptides
CC   by culture of the transformants; (5) preparation of avermectin aglycon
CC   or its derivatives by culture of transformed avermectin-producing
CC   microorganisms; and (6) oligonucleotides of 5-60 bases in length
CC   containing sense or antisense sequences from the avermectin aglycon
CC   synthase DNA. The enzymes are useful for the production of modified
CC   forms of avermectin and of the intermediates in its biosynthesis, for
CC   use as drugs, veterinary drugs and agrochemicals.
XX
SQ   Sequence 31422 BP; 4136 A; 10238 C; 11677 G; 5371 T; 0 other;

  Query Match          94.5%;  Score 873.4;  DB 21;  Length 31422;
  Best Local Similarity   96.7%;  Pred. No. 9.2e-120;
  Matches  908;  Conservative    0;  Mismatches   16;  Indels   15;  Gaps    1;

Qy          1 ACCGCATTCATCTGCTCCGGACAGGGCACCCAACGCCCCGGCATGGCCCACGGCCTCTAC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      22090 ACCGCATTCATCTGCTCCGGACAGGGCACCCAACGCCCCGGCATGGCCCACGGCCTCTAC 22149

Qy         61 CACACCCACCCCGTCTTCGCCGCCGCACTCAACGACATCTGCACCCACCTCGACCCCCAC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      22150 CACACCCACCCCGTCTTCGCCGCCGCACTCAACGACATCTGCACCCACCTCGACCCCCAC 22209

Qy        121 CTCGACCACCCCCTCCTCCCCCTCCTCACCCAAAACGACAACGACAA------------ 167
              |||||||||||||||||||||||||||||||  ||  ||||  | |
Db      22210 CTCGACCACCCCCTCCTCCCCCTCCTCACCCAGGACCCCAACACCCAGGACACCACCACC 22269

Qy        168 --CGAGGACGCGGCCGCACTGCTCCAGCAGACCCGCTACGCCCAGCCCGCCCTCTTCGCC 225
                ||| || ||||||||||||||||||||||||  ||||||||||||||||||||||||
Db      22270 CTCGAAGAAGCGGCCGCACTGCTCCAGCAGACCCCGTACGCCCAGCCCGCCCTCTTCGCC 22329

Qy        226 TTCCAGGTCGCCCTCCACCGCCTCCTCACCGACGGCTACCACATCACCCCCCACTACTAC 285
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      22330 TTCCAGGTCGCCCTCCACCGCCTCCTCACCGACGGCTACCACATCACCCCCCACTACTAC 22389

Qy        286 GCCGGACACTCCCTCGGCGAAATCACCGCCGCCCACCTCGCCGGCATCCTCACCCTCACC 345
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      22390 GCCGGACACTCCCTCGGCGAAATCACCGCCGCCCACCTCGCCGGCATCCTCACCCTCACC 22449

Qy        346 GACGCCACCACCCTCATCACCCAACGCGCCACCCTCATGCAAACCATGCCCCCCGGCACC 405
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      22450 GACGCCACCACCCTCATCACCCAACGCGCCACCCTCATGCAAACCATGCCCCCCGGCACC 22509

Qy        406 ATGACCACCCTCCACACCACCCCCCACCACATCACCCACCACCTCACCGCCCACGAAAAC 465
              |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db      22510 ATGACCACCCTCCACACCACCCCCCACCACATCACCCACCACATCACCGCCCACGAAAAC 22569

Qy        466 GACCTCGCCATCGCCGCCATCAACACCCCCACCTCCCTCGTCATCAGCGGCACCCCCCAC 525
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      22570 GACCTCGCCATCGCCGCCATCAACACCCCCACCTCCCTCGTCATCAGCGGCACCCCCCAC 22629

Qy        526 ACCGTCCAACACATCACCACCCTCTGCCAACAACAAGGCATCAAAACCAAAACCCTCCCC 585
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      22630 ACCGTCCAACACATCACCACCCTCTGCCAACAACAAGGCATCAAAACCAAAACCCTCCCC 22689

Qy        586 ACCAACCACGCCTTCCACTCCCCCCACACCAACCCCATCCTCAACCAACTCCACCAGCAC 645
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      22690 ACCAACCACGCCTTCCACTCCCCCCACACCAACCCCATCCTCAACCAACTCCACCAGCAC 22749

Qy        646 ACCCAAACCCTCACCTACCACCCACCCCACACCCCCCTCATCACCGCCAACACCCCACCC 705
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      22750 ACCCAAACCCTCACCTACCACCCACCCCACACCCCCCTCATCACCGCCAACACCCCACCC 22809

Qy        706 GACCAACTCCTCACCCCCCACTACTGGACCCAACAAGCCCGCAACACCGTCGACTACGCC 765
              |||||||||||||||||||||||||||||||||||||||||||||||||||||   |||
Db      22810 GACCAACTCCTCACCCCCCACTACTGGACCCAACAAGCCCGCAACACCGTCGACATAGCC 22869
```

```
Qy        766 ACCACCACCCAAACCCTCCACCAACACGGCGTCACCACCTACATCGAACTCGGACCCGAC 825
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      22870 ACCACCACCCAAACCCTCCACCAACACGGCGTCACCACCTACATCGAACTCGGACCCGAC 22929

Qy        826 AACACCCTCACCACCCTCACCCACCACAACCTCCCCAACCCCCCCACCACCACCCTCACC 885
              |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db      22930 AACACCCTCACCACCCTCACCCACCACAACCTCCCCAACACCCCCACCACCACCCTCACC 22989

Qy        886 CTCACCCACCCCCACCACCACCCCCAAACCCACCTCCTC 924
              |||||||||||||||||||||||||||||||||||||||
Db      22990 CTCACCCACCCCCACCACCACCCCCAAACCCACCTCCTC 23028


RESULT 6
AAH79278
ID   AAH79278 standard; DNA; 31422 BP.
XX
AC   AAH79278;
XX
DT   04-DEC-2001  (first entry)
XX
DE   Streptomyces avermitilis coding sequences SEQ ID NO: 2.
XX
KW   Avermectin aglycone synthase; AAS; avermectin derivative;
KW   drug production; vetinary drug; pesticide; ds.
XX
OS   Streptomyces avermitilis.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..14646
FT                   /*tag=  a
FT                   /product= "AAG65266"
FT   CDS             14824..31422
FT                   /*tag=  b
FT                   /product= "AAG65267"
XX
PN   WO200162939-A1.
XX
PD   30-AUG-2001.
XX
PF   23-FEB-2001; 2001WO-JP01381.
XX
PR   24-FEB-2000; 2000JP-0047405.
XX
PA   (KYOW ) KYOWA HAKKO KOGYO KK.
PA   (KITA ) KITASATO INST.
XX
PI   Endo H,  Yamaguchi H,  Kanda Y,  Hashimoto S,  Omura S,  Ikeda H;
XX
DR   WPI; 2001-582053/65.
DR   P-PSDB; AAG65266, AAG65267.
XX
PT   New modified avermectin aglycone synthase derived from Streptomyces
PT   avermectilis used in production of 22,23-dihydroavermectin B1a used in
PT   drugs and pesticides -
XX
PS   Disclosure; Page 103-149; 257pp; Japanese.
XX
CC   The present invention relates to the production of modified derivatives
CC   of avermectin aglycone synthase (AAS) derived from Streptomyces
CC   avermitilis. The activity of an acyl carrier protein (ACP),
CC   beta-ketoacyl carrier protein synthase (KS), acyltransferase (AT),
CC   beta-ketoacyl carrier protein reductase (KR), dehydratase (DH), enoyl
CC   reductase (ER) and/or thioesterase (TE) domain may be reduced or
CC   suppressed. The process can be used in the production of drugs, vetinary
CC   drugs and pesticides. The present sequence is a fragment of the S.
CC   avermitilis genome.
XX
SQ   Sequence 31422 BP; 4136 A; 10237 C; 11677 G; 5372 T; 0 other;

  Query Match             94.5%;  Score 873.4;  DB 22;  Length 31422;
  Best Local Similarity   96.7%;  Pred. No. 9.2e-120;
  Matches  908;  Conservative    0;  Mismatches   16;  Indels   15;  Gaps    1;

Qy          1 ACCGCATTCATCTGCTCCGGACAGGGCACCCAACGCCCCGGCATGGCCCACGGCCTCTAC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      22090 ACCGCATTCATCTGCTCCGGACAGGGCACCCAACGCCCCGGCATGGCCCACGGCCTCTAC 22149

Qy         61 CACACCCACCCCGTCTTCGCCGCCGCACTCAACGACATCTGCACCCACCTCGACCCCCAC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      22150 CACACCCACCCCGTCTTCGCCGCCGCACTCAACGACATCTGCACCCACCTCGACCCCCAC 22209

Qy        121 CTCGACCACCCCCTCCTCCCCCTCCTCACCCAAAACGACAACGACAA------------- 167
              |||||||||||||||||||||||||||||||||  ||  ||||  | |
Db      22210 CTCGACCACCCCCTCCTCCCCCTCCTCACCCAGGACCCCAACACCCAGGACACCACCACC 22269

Qy        168 --CGAGGACGCGGCCGCACTGCTCCAGCAGACCCGCTACGCCCAGCCCGCCCTCTTCGCC 225
                ||| || ||||||||||||||||||||||||  ||||||||||||||||||||||||
Db      22270 CTCGAAGAAGCGGCCGCACTGCTCCAGCAGACCCCGTACGCCCAGCCCGCCCTCTTCGCC 22329

Qy        226 TTCCAGGTCGCCCTCCACCGCCTCCTCACCGACGGCTACCACATCACCCCCCACTACTAC 285
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      22330 TTCCAGGTCGCCCTCCACCGCCTCCTCACCGACGGCTACCACATCACCCCCCACTACTAC 22389

Qy        286 GCCGGACACTCCCTCGGCGAAATCACCGCCGCCCACCTCGCCGGCATCCTCACCCTCACC 345
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      22390 GCCGGACACTCCCTCGGCGAAATCACCGCCGCCCACCTCGCCGGCATCCTCACCCTCACC 22449

Qy        346 GACGCCACCACCCTCATCACCCAACGCGCCACCCTCATGCAAACCATGCCCCCCGGCACC 405
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      22450 GACGCCACCACCCTCATCACCCAACGCGCCACCCTCATGCAAACCATGCCCCCCGGCACC 22509

Qy        406 ATGACCACCCTCCACACCACCCCCCACCACATCACCCACCACCTCACCGCCCACGAAAAC 465
              |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db      22510 ATGACCACCCTCCACACCACCCCCCACCACATCACCCACCACATCACCGCCCACGAAAAC 22569

Qy        466 GACCTCGCCATCGCCGCCATCAACACCCCCACCTCCCTCGTCATCAGCGGCACCCCCCAC 525
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      22570 GACCTCGCCATCGCCGCCATCAACACCCCCACCTCCCTCGTCATCAGCGGCACCCCCCAC 22629

Qy        526 ACCGTCCAACACATCACCACCCTCTGCCAACAACAAGGCATCAAAACCAAAACCCTCCCC 585
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      22630 ACCGTCCAACACATCACCACCCTCTGCCAACAACAAGGCATCAAAACCAAAACCCTCCCC 22689

Qy        586 ACCAACCACGCCTTCCACTCCCCCCACACCAACCCCATCCTCAACCAACTCCACCAGCAC 645
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      22690 ACCAACCACGCCTTCCACTCCCCCCACACCAACCCCATCCTCAACCAACTCCACCAGCAC 22749

Qy        646 ACCCAAACCCTCACCTACCACCCACCCCACACCCCCCTCATCACCGCCAACACCCCACCC 705
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      22750 ACCCAAACCCTCACCTACCACCCACCCCACACCCCCCTCATCACCGCCAACACCCCACCC 22809

Qy        706 GACCAACTCCTCACCCCCCACTACTGGACCCAACAAGCCCGCAACACCGTCGACTACGCC 765
              |||||||||||||||||||||||||||||||||||||||||||||||||||    |||
Db      22810 GACCAACTCCTCACCCCCCACTACTGGACCCAACAAGCCCGCAACACCGTCGACATAGCC 22869

Qy        766 ACCACCACCCAAACCCTCCACCAACACGGCGTCACCACCTACATCGAACTCGGACCCGAC 825
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      22870 ACCACCACCCAAACCCTCCACCAACACGGCGTCACCACCTACATCGAACTCGGACCCGAC 22929

Qy        826 AACACCCTCACCACCCTCACCCACCACAACCTCCCCAACCCCCCCACCACCACCCTCACC 885
              |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db      22930 AACACCCTCACCACCCTCACCCACCACAACCTCCCCAACACCCCCACCACCACCCTCACC 22989

Qy        886 CTCACCCACCCCCACCACCACCCCCAAACCCACCTCCTC 924
              |||||||||||||||||||||||||||||||||||||||
Db      22990 CTCACCCACCCCCACCACCACCCCCAAACCCACCTCCTC 23028


RESULT 7
AAF24892
ID   AAF24892 standard; DNA; 20394 BP.
XX
AC   AAF24892;
XX
DT   20-APR-2001  (first entry)
```

```
XX
DE   Pimaricin biosynthesis associated polyketide synthase gene.
XX
KW   Polyketide synthase; oxidative modification; metabolite; antibiotic;
KW   anticancer; pimaricin; ss.
XX
OS   Streptomyces natalensis.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..20394
FT                   /*tag= a
FT                   /product= "polyketide synthase"
XX
PN   WO200077222-A1.
XX
PD   21-DEC-2000.
XX
PF   14-JUN-2000; 2000WO-EP06227.
XX
PR   14-JUN-1999;   99EP-0201893.
XX
PA   (STAM ) DSM NV.
XX
PI   Martin JF,  Aparicio JF,  Colina AJ;
XX
DR   WPI; 2001-080693/09.
DR   P-PSDB; AAB31558.
XX
PT   New polynucleotides encoding enzymes involved in the biosynthesis of
PT   pimaricin, useful for modifying the biosynthesis of pimaricin and in
PT   the synthesis of new compounds -
XX
PS   Disclosure; Page 53-80; 116pp; English.
XX
CC   The present sequence encodes a polyketide synthase which is associated
CC   with the biosynthesis of pimaricin. The polyketide synthase polypeptide
CC   is useful for the oxidative modification of a methyl group of a suitable
CC   compound, e.g. a bioactive compound including a secondary metabolite,
CC   antibiotics and anticancer agents. Recombinant cells comprising the
CC   gene are useful for the production of pimaricin. The polyketide synthase
CC   polynucleotide may be over expressed in Streptomyces, leading to an
CC   increase in the biosynthesis of pimaricin, as a source of primers for
CC   amplification reaction and as probes.
XX
SQ   Sequence 20394 BP; 2611 A; 7486 C; 7397 G; 2900 T; 0 other;

  Query Match             32.7%;  Score 302.2;  DB 22;  Length 20394;
  Best Local Similarity   60.9%;  Pred. No. 2.9e-36;
  Matches  572;  Conservative    0;  Mismatches  343;  Indels   24;  Gaps.   4;

Qy          1 ACCGCATTCATCTGCTCCGGACAGGGCACCCAACGCCCCGGCATGGCCCACGGCCTCTAC 60
              | ||| ||| ||| | |||| |||||| |||| || || |||||||  | ||  ||  ||
Db       1687 AGCGCGTTCCTCTTCGCCGGGCAGGGCTCCCAGCGACCGGGCATGGGGCGCGAACTGCAC 1746

Qy         61 CACACCCACCCCGTCTTCGCCGCCGCACTCAACGACATCTGCACCCACCTCGACCCCCAC 120
                | | |||||||| |||||| | ||  || |||   | ||| || | || ||||| |||
Db       1747 GCCGCGCACCCCGTGTTCGCCACGGCCTTCGACGCGGTGTGCGCCGAACTGGACCCACAC 1806

Qy        121 CTCGACCACCCCCTCCTCCCCCTCCTCACCCAAAACGACAACGACAACGAGGACGCGGCC 180
              || ||||  || || | |  | || ||         || | | ||   |   | || ||||
Db       1807 CTGGACCGGCCGCTGCGCGACATCGTC------TTCGCCGAGGAAGGCAGCGCCGAGGCC 1860

Qy        181 GCACTGCTCCAGCAGACCCGCTACGCCCAGCCCGCCCTCTTCGCCTTCCAGGTCGCCCTC 240
              || |||||| | ||||||  |||| ||||| |||| |||||||| |   |   ||||||
Db       1861 GCCCTGCTCGACCAGACCGCCTACACCCAGGCCGCGCTCTTCGCCCTGGAAACCGCCCTG 1920

Qy        241 CACCGCCTCCTCACCGACGGCTACCACATCACCCCCCACTACTACGCCGGACACTCCCTC 300
                |||  |||  |   |||   ||    | |  | ||||  | |  ||||||||||||  ||
Db       1921 TTCCGGCTCGT---CGAATCCTGGGGCGTGGCACCCCGGTTCGTCGCCGGACACTCCATC 1977

Qy        301 GGCGAAATCACCGCCGCCCACCTCGCCGGCATCCTCACCCTCACCGACGCCACCACCCTC 360
              |||||  | |||||||||||| ||   ||| | || ||||||  |||||||  |    ||
Db       1978 GGCGAGCTGACCGCCGCCCACGTCAGTGGCGTGCTGACCCTCCACGACGCCGCACGGCTG 2037

Qy        361 ATCACCCAACGCGCCACCCTCATGCAAACCATGCCCCC---CGGCACCATGACCACCCTC 417
              || ||   |||| ||||||||||||  |  ||||| |   |||| |||||     |  ||
Db       2038 GTCGCCGCGCGCGGCACCCTCATGCAGGCGCTGCCCGCAGGCGGCGCCATGGTGGCGGTC 2097

Qy        418 CACACCACCCCCCACCACATCACCCACCACCTCACCGCCCACGAAAACGACCTCGCCATC 477
              ||  |||||    || | ||| | | |  ||| ||| |||||  || || ||||| ||
Db       2098 CAGGCCACCGAGGACGAGATCCGCGAGCGTCTCGCCGGCCACGAGGACCACGTCGCCCTC 2157

Qy        478 GCCGCCATCAACACCCCCACCTCCCTCGTCATCAGCGGCACCCCCCACACCGTCCAACAC 537
              || |||  ||||  |||  |||  |||||    |||  |   ||  |||||    |
Db       2158 GCGGCCGCCAACGGGCCCGATTCCACCGTCATTTCGGGCGACGAACAGGCCGTCACCGAG 2217

Qy        538 ATCACCACCCTCTGCCAACAACAAGGCATCAAAACCAAAACCCTCCCCACCAACCACGCC 597
              |||  |||  |||  |   ||| |||  |   |||||    ||  |   || |||||||
Db       2218 ATCGCGGCCCACTGGGAGGCACAGGGCCGCCGCACCAAGCGGCTGCGGGTCAGCCACGCC 2277

Qy        598 TTCCACTCCCCCCACACCAACCCCATCCTCAACCAACTCCACCAGCACACCCAAACCCTC 657
              |||||||||| ||||   ||  ||| ||  |  |  |||  |    | |||     ||
Db       2278 TTCCACTCCCCGCACATGGACGACATGCTGGAGGACTTCCGGCGCGTCGCCCGCGGTCTG 2337

Qy        658 ACCTACCACCCACCCCACACCCCC-----------CTCATCACCGCCAACACCCCACCC 705
              |||| |||| | |||| || ||||                | |  |  |||   | || |
Db       2338 ACCTTCCACGCCCCCCGCATCCCCGTGGTGTCCACGGTGACCGGCGCGCTCGCCACCGAA 2397

Qy        706 GACCAACTCCTCACCCCCCACTACTGGACCCAACAAGCCCGCAACACCGTCGACTACGCC 765
              ||| |||| | | | ||| ||||||||    |  || | |||| | ||||||  || |
Db       2398 GACGAACTGCGCTCGCCCGACTACTGGGTGCGGCAGGTCCGCGAAACCGTCCGCTTCTGT 2457

Qy        766 ACCACCACCCAAACCCTCCACCAACACGGCGTCACCACCTACATCGAACTCGGACCCGAC 825
              |  ||   |  |||||  |    | ||||||||||||  | | ||  |||| |||  |
Db       2458 GCTGCCGTGCGCACCCTTGAGGCCGAGGGCGTCACCACCTTCGTGGAGATCGGCACCGGC 2517

Qy        826 AACACCCTCACCACCCTCACCCACCACAACCTCCCCAACCCCCCCACCACCACCCTCACC 885
              |  ||||||| || |   |||  ||   ||  ||| || |         ||   |||  |
Db       2518 GGCGTCCTCACCCCCATGGTCCAGGACTGTCTGACCACCCTCGAAGAGCCCGTTCTCGTC 2577

Qy        886 CTCACCCACCCCCACCACCACCCCCAAACCCACCTCCTC 924
              |     | | |  | || |||| |||||  |   |||
Db       2578 CCTCTGCTGCGCACCGGCCGCCCCGAAACCGTCGCTCTC 2616


RESULT 8
AAD17184
ID   AAD17184 standard; DNA; 65140 BP.
XX
AC   AAD17184;
XX
DT   29-NOV-2001  (first entry)
XX
DE   Streptomyces noursei nys1 DNA of nystatin PKS gene cluster.
XX
KW   Polyketide synthase; PKS; macrolide; nystatin; PKS gene cluster;
KW   antifungal; antibiotic; nys1; ds.
XX
OS   Streptomyces noursei.
XX
FH   Key             Location/Qualifiers
FT   CDS             complement (1..1035)
FT                   /*tag= a
FT                   /product= "NysD2 partial protein"
FT                   /note= "CDS does not include stop codon"
FT   CDS             complement (1056..2576)
FT                   /*tag= b
FT                   /product= "NysD1 protein"
FT   CDS             2806..6906
FT                   /*tag= c
FT                   /product= "NysA protein"
FT   CDS             6952..16530
FT                   /*tag= d
FT                   /product= "NysB protein"
FT   CDS             16550..49840
FT                   /*tag= e
```

```
FT                   /product= "NysC protein"
FT   CDS             50260..51015
FT                   /*tag=  f
FT                   /product= "NysE protein"
FT   CDS             51405..54305
FT                   /*tag=  g
FT                   /product= "NysR1 protein"
FT   CDS             54329..57190
FT                   /*tag=  h
FT                   /product= "NysR2 protein"
FT                   /note= "CDS does not include start codon"
FT   CDS             57180..59963
FT                   /*tag=  i
FT                   /product= "NysR3 protein"
FT   CDS             60415..61047
FT                   /*tag=  j
FT                   /product= "NysR4 (short) protein"
FT                   /note= "CDS does not include start codon"
FT   CDS             61736..62497
FT                   /*tag=  k
FT                   /product= "NysR5 protein"
FT                   /note= "CDS does not include start codon"
FT   CDS             complement (62551..63615)
FT                   /*tag=  l
FT                   /product= "ORF2 protein"
FT                   /note= "CDS does not include start codon"
FT   CDS             63765..64961
FT                   /*tag=  m
FT                   /product= "ORF1 protein"
XX
PN   WO200159126-A2.
XX
PD   16-AUG-2001.
XX
PF   08-FEB-2001; 2001WO-GB00509.
XX
PR   08-FEB-2000; 2000GB-0002840.
PR   10-APR-2000; 2000GB-0008786.
PR   14-APR-2000; 2000GB-0009387.
XX
PA   (UYNO-) UNIV NORGES TEKNISK NATURVITENSKAPELIGE.
PA   (SNTF ) SINTEF STIFTELSEN IND TEK FORSK.
PA   (ALPH-) ALPHARMA AS.
PA   (SINV-) SINVENT AS.
PA   (DZIE/) DZIEGLEWSKA H.
PA   (ZOTC/) ZOTCHEV S B.
PA   (SEKU/) SEKUROVA O N.
PA   (FJAE/) FJAERVIK E.
PA   (BRAU/) BRAUTASET T.
PA   (STRO/) STROM A R.
XX
PI   Zotchev SB,  Sekurova ON,  Fjaervik E,  Brautaset T,  Strom AR;
PI   Valla S,  Ellingsen TE,  Sletta H,  Gulliksen O;
XX
DR   WPI; 2001-557614/62.
DR   P-PSDB; AAE10125, AAE10126, AAE10127, AAE10128, AAE10129, AAE10130,
DR   AAE10131, AAE10132, AAE10133, AAE10134, AAE10135, AAE10136, AAE10137.
XX
PT   New nystatin polyketide synthase polynucleotides and polypeptides,
PT   useful as antibiotics and antifungals -
XX
PS   Claim 2; Page 116-151; 266pp; English.
XX
CC   The present invention relates to the cloning and sequencing of the gene
CC   cluster encoding a modular type I polyketide synthase (PKS) enzyme
CC   involved in the biosynthesis of the macrolide antibiotic nystatin.
CC   The nystatin PKS is useful as antifungal antibiotics. The present
CC   sequence is a Streptomyces noursei nys1 DNA of nystatin PKS gene cluster.
XX
SQ   Sequence 65140 BP; 8270 A; 25171 C; 22273 G; 9426 T; 0 other;

  Query Match            32.5%;  Score 300.6;  DB 22;  Length 65140;
  Best Local Similarity  62.1%;  Pred. No. 4.4e-36;

    Matches  581;  Conservative    0;  Mismatches  309;  Indels   45;  Gaps    5;

Qy          1 ACCGCATTCATCTGCTCCGGACAGGGCACCCAACGCCCCGGCATGGCCCACGGCCTCTAC 60
              ||||| ||| ||| |||||| |||||| ||||||| | |||||||| || || |||||||
Db      23408 ACCGCCTTCCTCTTCTCCGGCCAGGGCTCCCAACGGCTCGGCATGGGCCGCGTCCTCTAC 23467

Qy         61 CACACCCACCCCGTCTTCGCCGCCGCACTCAACGACATCTGCACCCACCTCGACCCCCAC 120
              |      ||||| |||||||| || ||| ||  | ||  |||| |||||||  |  |
Db      23468 GAGCGGTTCCCCGCCTTCGCCGAAGCCCTCGACACCGTCCTCACCGCCCTCGACGCGGAA 23527

Qy        121 CTCGACCACCCCCTCCTCCCCCTCCTCACCCAAAACGACAACGACAACGAGGACGCGGCC 180
              |||| ||||||||||                    |||||  |  |    |  || |  ||
Db      23528 CTCGGCCACCCCCTCC------------------GCGACATCATCTGGGGCGAGGACGCT 23569

Qy        181 GCACTGCTCCAGCAGACCCGCTACGCCCAGCCCGCCCTCTTCGCCTTCCAGGTCGCCCTC 240
              |||  || | | |||| ||||| |||| |||||||| |||||| || |||| || |||
Db      23570 CAACTCGTCGACCGGACCGGCTACACCCAACCCGCCCTGTTCGCCATCGAGGTGGCACTC 23629

Qy        241 CACCGCCTCCTCACCGACGGCTACCACATCACCCCCCACTACTACGCCGGACACTCCCTC 300
              |||||||||     || | ||     |||||| ||  ||| |   |||| ||||| || ||
Db      23630 TTCCGCCTCCT---GGAAGCCTGGGGCATCACACCGGACTTCGTGGCCGGCCACTCCATC 23686

Qy        301 GGCGAAATCACCGCCGCCCACCTCGCCGGCATCCTCACCCTCACCGACGCCACCACCCTC 360
              ||||| ||| ||||||| ||| |||||||| | ||| |||||  |||||||   |  ||||
Db      23687 GGCGAGATCGCCGCCGCACACGTCGCCGGCGTGCTCTCCCTCGGCGACGCCTGCCGCCTC 23746

Qy        361 ATCACCCAACGCGCCACCCTCATGCAAACCATGCCC---CCCGGCACCATGACCACCCTC 417
              ||       ||||||   || |||||  |  |||||     |||| | |||| | || ||
Db      23747 GTCGTGGCCCGCCGCCGTGCTGATGCAGTCGCTGCCCGAAGGCGGCGCGATGATCGCCGTC 23806

Qy        418 CACACCACCCCCCACCACATCACCCACCACCTCACCGCCCACGAAAACGACCTCGCCATC 477
              ||  |||||    || |  ||   | || ||||||||          ||||| || | |||
Db      23807 CAGGCCACCGAGGACGAGGTCCTGCCCCTCCTCACCG---------ACGACGTCTCGATC 23857

Qy        478 GCCGCCATCAACACCCCCACCTCCCTCGTCATCAGCGGCACCCCCCCACACCGTCCAACAC 537
              |||||| |||||| ||| |||||| ||||| ||  ||||  |    || ||  ||    |
Db      23858 GCCGCCGTCAACAGCCCGACCTCCGTCGTCGTCTCCGGCTACGAGAACGCCACCCTCGCC 23917

Qy        538 ATCACCACCCTCTGCCAACAACAAGGCATCAAAACCAAAACCCTCCCCACCAACCACGCC 597
              || ||   | || |    | || |||  |   ||||     || | |  || |||||||
Db      23918 GTCGCCCGGCACTTCGCCGACCAGGGCCGCCGCACCACGCGGCTGCGCGTCAGCCACGCC 23977

Qy        598 TTCCACTCCCCCCACACCAACCCCATCCTCAACCAACTCCACCAGCACACCCAAACCCTC 657
              |||||||| || |  |    || || ||| || |  ||| |     |  | | ||||
Db      23978 TTCCACTCGCCGCTGATGGCGCCGATGCTCGACGACTTCCGCGCCGTCGTCGAGAGCCTC 24037

Qy        658 ACCTACCACCCACCCCCACACCCCCCTCATCACCGCCAACA-----------CCCCACCC 705
              |||| |  | | |||   |||||| || || ||  |   |             ||||  ||
Db      24038 ACCTTCACCGCCCCCACGACCCCCGTCGTCTCCAACCTGACCGGCGAACTGGCCCCGGCC 24097

Qy        706 GACCAACTCCTCACCCCCCACTACTGGACCCAACAAGCCCGCAACACCGTCGACTACGCC 765
              ||    |||  | |  || ||||||||  ||  || | |||| |  | |||  || ||||
Db      24098 GAGGCGCTCTGCTCGGCCGACTACTGGGTCCGGCACGTCCGCGAGGCGGTCCGCTTCGCC 24157

Qy        766 ACCACCACCCAAACCCTCCACCAACACGGCGTCACCACCTACATCGAACTCGGACCCGAC 825
              |  || ||  ||||||  | | | ||||||||||||||| | ||||||||| ||||||
Db      24158 GACGGCATCCGCACCCTCGCCGACCGCGGCGTCACCACCTTCGTCGAACTCGGCCCCGAC 24217

Qy        826 AACACCCTCACCACCCTCACCCACCACAACCTCCCCAACCCCCCCACCACCACCCTCACC 885
              | |   ||  || || |  |||| |  |  |||| |  | ||  ||||| ||    |
Db      24218 AGCGTGCTGTCCGCCATGGCCCAGGAGTCCGCCCCCGAAGGCGCCGGCACCATCCCGCTC 24277

Qy        886 CTCACCCACCCCCACCACCACCCCCAAACCCACCT 920
              ||  || |  ||  | | |    ||  ||  |||
Db      24278 CTGCGCCGCGACCGGCCCGAGGAACAGGCCGTCCT 24312


RESULT 9
AAD17186
ID   AAD17186 standard; DNA; 125401 BP.
XX
AC   AAD17186;
XX
```

```
DT   29-NOV-2001  (first entry)
XX
DE   Streptomyces noursei nystatin PKS gene cluster DNA.
XX
KW   Polyketide synthase; PKS; macrolide; nystatin; PKS gene cluster;
KW   antifungal; antibiotic; ds.
XX
OS   Streptomyces noursei.
XX
FH   Key             Location/Qualifiers
FT   CDS             6337..34771
FT                   /*tag= a
FT                   /product= "NysI complete protein"
FT   CDS             34792..51099
FT                   /*tag= b
FT                   /product= "NysJ protein"
FT   CDS             51155..57355
FT                   /*tag= c
FT                   /product= "NysK protein"
FT   CDS             57503..58687
FT                   /*tag= d
FT                   /product= "NysL protein"
FT   CDS             complement (58786..58980)
FT                   /*tag= e
FT                   /product= "NysM protein"
FT                   /note= "CDS does not include start codon"
FT   CDS             complement (59045..60241)
FT                   /*tag= f
FT                   /product= "NysN protein"
FT                   /note= "CDS does not include start codon"
FT   CDS             complement (60238..61296)
FT                   /*tag= g
FT                   /product= "NysD2 complete protein"
FT   CDS             120628..121308
FT                   /*tag= h
FT                   /product= "NysR4 (long) protein"
XX
PN   WO200159126-A2.
XX
PD   16-AUG-2001.
XX
PF   08-FEB-2001; 2001WO-GB00509.
XX
PR   08-FEB-2000; 2000GB-0002840.
PR   10-APR-2000; 2000GB-0008786.
PR   14-APR-2000; 2000GB-0009387.
XX
PA   (UYNO-) UNIV NORGES TEKNISK NATURVITENSKAPELIGE.
PA   (SNTF ) SINTEF STIFTELSEN IND TEK FORSK.
PA   (ALPH-) ALPHARMA AS.
PA   (SINV-) SINVENT AS.
PA   (DZIE/) DZIEGLEWSKA H.
PA   (ZOTC/) ZOTCHEV S B.
PA   (SEKU/) SEKUROVA O N.
PA   (FJAE/) FJAERVIK E.
PA   (BRAU/) BRAUTASET T.
PA   (STRO/) STROM A R.
XX
PI   Zotchev SB,  Sekurova ON,  Fjaervik E,  Brautaset T,  Strom AR;
PI   Valla S,  Ellingsen TE,  Sletta H,  Gulliksen O;
XX
DR   WPI; 2001-557614/62.
DR   P-PSDB; AAE10143, AAE10144, AAE10145, AAE10146, AAE10147, AAE10148,
DR   AAE10149, AAE10150.
XX
PT   New nystatin polyketide synthase polynucleotides and polypeptides,
PT   useful as antibiotics and antifungals -
XX
PS   Claim 1; Page 188-254; 266pp; English.
XX
CC   The present invention relates to the cloning and sequencing of the gene
CC   cluster encoding a modular type I polyketide synthase (PKS) enzyme
CC   involved in the biosynthesis of the macrolide antibiotic nystatin.
CC   The nystatin PKS is useful as antifungal antibiotics. The present
CC   sequence is a Streptomyces noursei nystatin PKS gene cluster DNA.
XX
SQ   Sequence 125401 BP; 15664 A; 49692 C; 42871 G; 17174 T; 0 other;

  Query Match           32.5%;  Score 300.6;  DB 22;  Length 125401;
  Best Local Similarity  62.1%;  Pred. No. 4.2e-36;
  Matches  581;  Conservative    0;  Mismatches  309;  Indels   45;  Gaps    5;

Qy          1 ACCGCATTCATCTGCTCCGGACAGGGCACCCAACGCCCCGGCATGGCCCACGGCCTCTAC 60
              ||||| ||| ||| |||||| |||||| ||||||| | |||||||| || || |||||||
Db      83669 ACCGCCTTCCTCTTCTCCGGCCAGGGCTCCCAACGGCTCGGCATGGGCCGCGTCCTCTAC 83728

Qy         61 CACACCCACCCCGTCTTCGCCGCCGCACTCAACGACATCTGCACCCACCTCGACCCCCAC 120
              |     ||||| |||||||| || ||| ||  | ||  |||| ||||||| |  |
Db      83729 GAGCGGTTCCCCGCCTTCGCCGAAGCCCTCGACACCGTCCTCACCGCCCTCGACGCGGAA 83788

Qy        121 CTCGACCACCCCCTCCTCCCCCTCCTCACCCAAAACGACAACGACAACGAGGACGCGGCC 180
              |||| ||||||||||                    ||||| |  |  |  || |  ||
Db      83789 CTCGGCCACCCCCTCC-----------------GCGACATCATCTGGGGCGAGGACGCT 83830

Qy        181 GCACTGCTCCAGCAGACCCGCTACGCCCAGCCCGCCCTCTTCGCCTTCCAGGTCGCCCTC 240
              |||  || | | |||| ||||| |||| |||||||| |||||| || |||| || |||
Db      83831 CAACTCGTCGACCGGACCGGCTACACCCAACCCGCCCTGTTCGCCATCGAGGTGGCACTC 83890

Qy        241 CACCGCCTCCTCACCGACGGCTACCACATCACCCCCCACTACTACGCCGGACACTCCCTC 300
              |||||||||    || | ||    |||||| ||  ||| |  ||||| |||||| ||
Db      83891 TTCCGCCTCCT---GGAAGCCTGGGGCATCACACCGGACTTCGTGGCCGGCCACTCCATC 83947

Qy        301 GGCGAAATCACCGCCGCCCACCTCGCCGGCATCCTCACCCTCACCGACGCCACCACCCTC 360
              ||||| ||| ||||||| ||| |||||||| | ||| ||||  |||||||  |  ||||
Db      83948 GGCGAGATCGCCGCCGCACACGTCGCCGGCGTGCTCTCCCTCGGCGACGCCTGCCGCCTC 84007

Qy        361 ATCACCCAACGCGCCACCCTCATGCAAACCATGCCC---CCCGGCACCATGACCACCCTC 417
              ||      ||||||   || |||||  | |||||     |||| | |||| | || ||
Db      84008 GTCGTGGCCCGCGCCGTGCTGATGCAGTCGCTGCCCGAAGGCGGCGCGATGATCGCCGTC 84067

Qy        418 CACACCACCCCCCACCACATCACCCACCACCTCACCGCCCACGAAAACGACCTCGCCATC 477
              ||  |||||  .  || |  ||   | || |||||||||          |||||  || | |||
Db      84068 CAGGCCACCGAGGACGAGGTCCTGCCCCTCCTCACCG---------ACGACGTCTCGATC 84118

Qy        478 GCCGCCATCAACACCCCCACCTCCCTCGTCATCAGCGGCACCCCCCCACACCGTCCAACAC 537
              |||||| |||||| ||| ||||| ||||| ||  ||||  |    ||  ||  ||     |
Db      84119 GCCGCCGTCAACAGCCCGACCTCCGTCGTCGTCTCCGGCTACGAGAACGCCACCCTCGCC 84178

Qy        538 ATCACCACCCTCTGCCAACAACAAGGCATCAAAACCAAAACCCTCCCCACCAACCACGCC 597
              || ||   | || |     | || ||| |  ||||      || | |  || |||||||
Db      84179 GTCGCCCGGCACTTCGCCGACCAGGGCCGCCGCACCACGCGGCTGCGCGTCAGCCACGCC 84238

Qy        598 TTCCACTCCCCCCACACCAACCCCATCCTCAACCAACTCCACCAGCACACCCAAACCCTC 657
              |||||||| || |   |     || || ||| || |  ||| |     |  | | | ||||
Db      84239 TTCCACTCGCCGCTGATGGCGCCGATGCTCGACGACTTCCGCGCCGTCGTCGAGAGCCTC 84298

Qy        658 ACCTACCACCCACCCCCACACCCCCCTCATCACCGCCAACA-----------CCCCACCC 705
              |||| |  | | |||   |||||| || || ||  |  |                ||||  ||
Db      84299 ACCTTCACCGCCCCCCACGACCCCCGTCGTCTCCAACCTGACCGGCGAACTGGCCCCGGCC 84358

Qy        706 GACCAACTCCTCACCCCCCACTACTGGACCCAACAAGCCCGCAACACCGTCGACTACGCC 765
              ||    |||  | |  || |||||||  ||  || | |||| |  | |||  || ||||
Db      84359 GAGGCGCTCTGCTCGGCCGACTACTGGGTCCGGCACGTCCGCGAGGCGGTCCGCTTCGCC 84418

Qy        766 ACCACCACCCAAACCCTCCACCAACACGGCGTCACCACCTACATCGAACTCGGACCCGAC 825
              |  || ||  |||||  | | | ||||||||||||||| | ||||||||||| ||||||
Db      84419 GACGGCATCCGCACCCTCGCCGACCGCGGCGTCACCACCTTCGTCGAACTCGGCCCCGAC 84478

Qy        826 AACACCCTCACCACCCTCACCCACCACAACCTCCCCAACCCCCCCACCACCACCCTCACC 885
              | |   ||  || || |  |||| |  |  |||| |  | ||  ||||| ||   |
Db      84479 AGCGTGCTGTCCGCCATGGCCCAGGAGTCCGCCCCCGAAGGCGCCGGCACCATCCCGCTC 84538

Qy        886 CTCACCCACCCCCACCACCACCCCCAAACCCACCT 920
              ||  || |  ||  | | |    ||  ||  |||
Db      84539 CTGCGCCGCGACCGGCCCGAGGAACAGGCCGTCCT 84573
```

```
RESULT 10
AAD17185
ID   AAD17185 standard; DNA; 27541 BP.
XX
AC   AAD17185;
XX
DT   29-NOV-2001  (first entry)
XX
DE   Streptomyces noursei nys2 DNA of nystatin PKS gene cluster.
XX
KW   Polyketide synthase; PKS; macrolide; nystatin; PKS gene cluster;
KW   antifungal; antibiotic; nys2; ds.
XX
OS   Streptomyces noursei.
XX
FH   Key             Location/Qualifiers
FT   CDS             complement (454..1191)
FT                   /*tag=  a
FT                   /product= "NysF protein"
FT                   /note= "CDS does not include start codon"
FT   CDS             complement (1275..3092)
FT                   /*tag=  b
FT                   /product= "NysG protein"
FT   CDS             complement (3070..4824)
FT                   /*tag=  c
FT                   /note= "CDS does not include start codon"
FT                   /product= "NysH protein"
FT   CDS             5122..6156
FT                   /*tag=  d
FT                   /product= "NysD3 protein"
FT   CDS             6338..27541
FT                   /*tag=  e
FT                   /product= "NysI partial protein"
FT                   /note= "CDS does not include stop codon"
XX
PN   WO200159126-A2.
XX
PD   16-AUG-2001.
XX
PF   08-FEB-2001; 2001WO-GB00509.
XX
PR   08-FEB-2000; 2000GB-0002840.
PR   10-APR-2000; 2000GB-0008786.
PR   14-APR-2000; 2000GB-0009387.
XX
PA   (UYNO-) UNIV NORGES TEKNISK NATURVITENSKAPELIGE.
PA   (SNTF ) SINTEF STIFTELSEN IND TEK FORSK.
PA   (ALPH-) ALPHARMA AS.
PA   (SINV-) SINVENT AS.
PA   (DZIE/) DZIEGLEWSKA H.
PA   (ZOTC/) ZOTCHEV S B.
PA   (SEKU/) SEKUROVA O N.
PA   (FJAE/) FJAERVIK E.
PA   (BRAU/) BRAUTASET T.
PA   (STRO/) STROM A R.
XX
PI   Zotchev SB,  Sekurova ON,  Fjaervik E,  Brautaset T,  Strom AR;
PI   Valla S,  Ellingsen TE,  Sletta H,  Gulliksen O;
XX
DR   WPI; 2001-557614/62.
DR   P-PSDB; AAE10138, AAE10139, AAE10140, AAE10141, AAE10142.
XX
PT   New nystatin polyketide synthase polynucleotides and polypeptides,
PT   useful as antibiotics and antifungals -
XX
PS   Claim 2; Page 151-166; 266pp; English.
XX
CC   The present invention relates to the cloning and sequencing of the gene
CC   cluster encoding a modular type I polyketide synthase (PKS) enzyme
CC   involved in the biosynthesis of the macrolide antibiotic nystatin.
CC   The nystatin PKS is useful as antifungal antibiotics. The present
CC   sequence is a Streptomyces noursei nys2 DNA of nystatin PKS gene cluster.
XX
SQ   Sequence 27541 BP; 3517 A; 10766 C; 9529 G; 3729 T; 0 other;

  Query Match           29.6%;  Score 273.4;  DB 22;  Length 27541;
  Best Local Similarity   59.7%;  Pred. No. 4.5e-32;
  Matches  551;  Conservative    0;  Mismatches  336;  Indels   36;  Gaps    4;

Qy          3 CGCATTCATCTGCTCCGGACAGGGCACCCAACGCCCCGGCATGGCCCACGGCCTCTACCA 62
              |||  || | | | |||| |||||| |||||||| ||||  ||  |  |  |||  |
Db       8041 CGCCGTCCTGTTCACCGGCCAGGGCGCCCAACGCCTCGGCGCGGGACGGGAACTCGCCGC 8100

Qy         63 CACCCACCCCGTCTTCGCCGCCGCACTCAACGACATCTGCACCCACCTCGACCCCCACCT 122
              |     ||| |  |||||| |||| ||| |||      .|  |  | ||  ||| |||||
Db       8101 CCGTTTCCCGGCGTTCGCCACCGCCCTCGACGCGGCGCTCGACGCCTTCACCCCGCACCT 8160

Qy        123 CGACCACCCCCTCCTCCCCCTCCTCACCCAAAACGACAACGACAACGAGGACGCGGCCGC 182
              ||||| |||||| | |   ||||       | ||||                    |||||
Db       8161 CGACCGCCCCCTGCGCGAGGTCCTGTGGGGCACCGAC----------------GCCGC 8202

Qy        183 ACTGCTCCAGCAGACCCGCTACGCCCAGCCCGCCCTCTTCGCCTTCCAGGTCGCCCTCCA 242
              |||||| | |  |||   |||||||||| |||||||||||| || | || || ||| |
Db       8203 CCTGCTCGACCGCACCGCATACGCCCAGCCGGCCCTCTTCGCCGTCGAAGTGGCGCTCTA 8262

Qy        243 CCGCCTCCTCACCGACGGCTACCACATCACCCCCCACTACTACGCCGGACACTCCCTCGG 302
              |||    || | |||    | |  | ||  |||| || ||  |||||| |||||| ||||
Db       8263 CCG---GCTGATCGAATCGTTCGGCGTCCGCCCCGACCACCTCGCCGGCCACTCCGTCGG 8319

Qy        303 CGAAATCACCGCCGCCCACCTCGCCGGCATCCTCACCCTCACCGACGCCACCACCCTCAT 362
              ||| |||  |||||| |||||||||| ||||| |||| |||||||| |||||||| |
Db       8320 CGAGATCGTCGCCGCGCACCTCGCCGGGGTCCTCTCCCTGGCCGACGCCGCCACCCTCGT 8379

Qy        363 CACCCAACGCGCCACCCTCATGCAAACCATGCCCCCCGGC---ACCATGACCACCCTCCA 419
              | ||   ||||    ||| |||||  |  |||||  ||||    | |||| | || ||||
Db       8380 CGCCGCCCGCGGTCGCCTGATGCAGGCGCTGCCCGACGGCGGGGCGATGATCGCCGTCCA 8439

Qy        420 CACCACCCCCCACCACATCACCCACCACCTCACCGCCCACGAAAACGACCTCGCCATCGC 479
              |  |      | || || |||  |  ||| |||  |||||  || |  |||| |||||
Db       8440 GGCGTCGGAAGCCGACGTCGCCCCGCTGCTCGCCGGGCACGAGGACCAGGTCGCGATCGC 8499

Qy        480 CGCCATCAACACCCCCACCTCCCTCGTCATCAGCGGCACCCCCCACACCGTCCAACACAT 539
              |||| |||||  |||| || || ||||| |   |||| ||    |||||||         |
Db       8500 CGCCGTCAACGGCCCCTCCGCCGTCGTCCTGTCCGGCGCCGAAGCCACCGTCACCGCGCT 8559

Qy        540 CACCACCCTCTGCCAACAACAAGGCATCAAAACCAAAACCCTCCCCACCAACCACGCCTT 599
              | ||   |   |       | |||  ||| |||       || | |  |  ||||||||
Db       8560 CGCCGAACAGCTCGCCGCCGACGGCCGCAAGACCCGCCGGCTGCGCGTCTCGCACGCCTT 8619

Qy        600 CCACTCCCCCCACACCAACCCCATCCTCAACCAACTCCACCAGCACACCCAAACCCTCAC 659
              |||||| || | ||   | || || ||| ||     ||| |    |  | ||  ||||||
Db       8620 CCACTCGCCGCTCATGGAGCCGATGCTCGACGCCTTCCGCGCCGTCGTCGAAGACCTCAC 8679

Qy        660 CTACCACCCACCCCACACCCCCCTCATCACCGCC----------AACACCCCACCCGA 707
                 ||| || || | |   ||  || || || |           | || |   ||
Db       8680 GCTCCAGCCGCCGCTCCTGCCGGTCGTCTCCAACCTGACCGGCAAGCCCGCCACCGTCGC 8739

Qy        708 CCAACTCCTCACCCCCCACTACTGGACCCAACAAGCCCGCAACACCGTCGACTACGCCAC 767
              |||||||  | || || ||||||||  | | || | |||  || |||||  || ||||
Db       8740 CCAACTCACCTCCGCCGACTACTGGGTCGACCACGTCCGGCACGCCGTCCGCTTCGCCGA 8799

Qy        768 CACCACCCAAACCCTCCACCAACACGGCGTCACCACCTACATCGAACTCGGACCCGACAA 827
              |  || | |   |||  ||  |||| |  |||| ||| | |||||||||| ||||||
Db       8800 CGGCATCGACTGGCTCGCCCGGCACGACACCACCGCCTTCCTCGAACTCGGCCCCGACGG 8859

Qy        828 CACCCTCACCACCCTCACCCACCACAACCTCCCCAACCCCCCCACCACCACCCTCACCCT 887
              |   ||  || || |  |||| ||  |||   |  | || | |  | || |||||||
Db       8860 CGTGCTGTCCGCCATGGCCCAGGACTGCCTGGACGCCGCCGACGCAGACGCCGTCACCCT 8919

Qy        888 CACCCACCCCCACCACCACCCCC 910
              | ||  ||  | |  |   | ||
Db       8920 CCCCGCCCTGCGCGCCGGGCGCC 8942


RESULT 11
```

```
AAT06769
ID   AAT06769 standard; DNA; 28598 BP.
XX
AC   AAT06769;
XX
DT   15-OCT-1996  (first entry)
XX
DE   Sorangium cellulosum soraphen gene cluster.
XX
KW   Antipathogenic substance; soraphen; phenylpyrrole; antibiotic;
KW   fungicide; pesticide; myxobacterium; ss.
XX
OS   Sorangium cellulosum.
XX
PN   WO9533818-A2.
XX
PD   14-DEC-1995.
XX
PF   30-MAY-1995;   95WO-IB00414.
XX
PR   08-JUN-1994;   94US-0258261.
XX
PA   (CIBA ) CIBA GEIGY AG.
XX
PI   Beck JJ,  Gaffney TD,  Hammer PE,  Hill DS,  Lam ST;
PI   Ligon J,  Ryals JA, Schupp T,  Uknes SJ;
XX
DR   WPI; 1996-040226/04.
XX
PT   New genes for biosynthesis of anti-pathogenic substances - pref.
PT   pyrrolnitrin and soraphen, useful for disease control in plants
XX
PS   Claim 3; Page 140-158; 190pp; English.
XX
CC   This is the soraphen gene cluster from S. cellulosum.  The gene
CC   cluster may be expressed recombinantly to produce soraphen, or
CC   expressed in a transgenic plant for disease-resistance.
XX
SQ   Sequence 28598 BP; 4031 A; 11703 C; 8263 G; 4601 T; 0 other;

  Query Match             28.8%;  Score 266.2;  DB 17;  Length 28598;
  Best Local Similarity   59.0%;  Pred. No. 5.1e-31;
  Matches  542;  Conservative    0;  Mismatches  353;  Indels   24;  Gaps    4;

Qy          3 CGCATTCATCTGCTCCGGACAGGGCACCCAACGCCCCGGCATGGCCCACGGCCTCTACCA 62
              ||| || ||| | | || || || | ||| || |||  ||||| || ||||||||||| |
Db       5250 CGCCGTCCTCTTCACGGGGCAAGGAAGCCAGCGGCCCACCATGGGCCGCGGCCTCTACGA 5309

Qy         63 CACCCACCCCGTCTTCGCCGCCGCACTCAACGACATCTGCACCCACCTCGACCCCCACCT 122
              |     ||||||||||    | ||| ||| ||  | || || |||||||||||| | | ||
Db       5310 CGTTTTCCCCGTCTTCCGGGACGCCCTCGACACCGTCGGCGCCCACCTCGACCGCGAGCT 5369

Qy        123 CGACCACCCCCTCCTCCCCCTCCTCACCCAAAACGACAACGACAACGAGGACGCGGCCGC 182
              ||||| |||||| | |  | |||||         ||     ||||  |    ||   ||||||
Db       5370 CGACCGCCCCCTGCGCGACGTCCTC------TTCGCTCCCGACGGCTCCGAGCAGGCCGC 5423

Qy        183 ACTGCTCCAGCAGACCCGCTACGCCCAGCCCGCCCTCTTCGCCTTCCAGGTCGCCCTCCA 242
              |  ||| |||| |||  || | ||||||| ||||| || ||| || | |||||||||
Db       5424 GCGCCTCGAGCAAACCGCCTTCACCCAGCCGGCCCTGTTTGCCCTCGAAGTCGCCCTCTT 5483

Qy        243 CCGCCTCCTCACCGACGGCTACCACATCACCCCCCACTACTACGCCGGACACTCCCTCGG 302
              |  || ||     |   || |    | |  |||     |  |  |||||||||| | ||
Db       5484 TCAGCTTCT---ACAATCCTTCGGTCTGAAGCCCGCTCTCCTCCTCGGACACTCCATTGG 5540

Qy        303 CGAAATCACCGCCGCCCACCTCGCCGGCATCCTCACCCTCACCGACGCCACCACCCTCAT 362
              |||  ||  |||||||||| |||||||| |||| | |||   |||| |  |||||| |
Db       5541 CGAGCTCGTCGCCGCCCACGTCGCCGGCGTCCTTTCTCTCCAGGACGGCTGCACCCTCGT 5600

Qy        363 CACCCAACGCGCCACCCTCATGCAAACCATGCCCC---CCGGCACCATGACCACCCTCCA 419
              | ||   ||||| |  |||||||||  |  | || |    |||| |||||  ||||||||
Db       5601 CGCCGCCCGCGCAAAGCTCATGCAAGCGCTCCCACAAGGCGGCGCCATGGTCACCCTCCG 5660

Qy        420 CACCACCCCCCACCACATCACCCACCACCTCACCGCCCACGAAAACGACCTCGCCATCGC 479
              || ||    |  |  ||  | |||  |||    || |||||  |         | ||||
Db       5661 AGCCTCCGAGGAGGAAGTCCGCGACCTTCTCCAGCCCTACGAAGGCCGAGCTAGCCTCGC 5720

Qy        480 CGCCATCAACACCCCCACCTCCCTCGTCATCAGCGGCACCCCCCCACACCGTCCAACACAT 539
              |||| ||||    ||   ||||  |||| ||   |||        || | ||     | ||
Db       5721 CGCCCTCAATGGGCCTCTCTCCACCGTCGTCGCTGGCGATGAAGACGCGGTGGTGGAGAT 5780

Qy        540 CACCACCCTCTGCCAACAACAAGGCATCAAAACCAAAACCCTCCCCACCAACCACGCCTT 599
              | ||  ||     | ||    |  ||     || |||| |  ||| | |  || |||||||||
Db       5781 CGCCCGCCAGGCCGAAGCCCTCGGACGAAAGACCACACGCCTGCGCGTCAGCCACGCCTT 5840

Qy        600 CCACTCCCCCCACACCAACCCCATCCTCAACCAACTCCACCAGCACACCCAAACCCTCAC 659
              ||| ||||| ||||    ||   || ||| || |  ||| ||     | |||| | ||||||
Db       5841 CCATTCCCCGCACATGGACGGAATGCTCGACGACTTCCGCCGCGTCGCCCAGAGCCTCAC 5900

Qy        660 CTACCACCCACCCCACACCCCCCTCATCAC------------CGCCAACACCCCACCCGA 707
              |||||| ||  | | || |||| ||||| |             |||   | || |   | |
Db       5901 CTACCATCCCGCACGCATCCCCATCATCTCCAACGTCACCGGCGCGCGCGCCACGGACCA 5960

Qy        708 CCAACTCCTCACCCCCCACTACTGGACCCAACAAGCCCGCAACACCGTCGACTACGCCAC 767
              | | |||  | | ||| ||||||||  ||  || |  ||| ||||||||  || |  |
Db       5961 CGAGCTCGCCTCGCCCGACTACTGGGTCCGCCACGTTCGCCACACCGTCCGCTTCCTCGA 6020

Qy        768 CACCACCCAAACCCTCCACCAACACGGCGTCACCACCTACATCGAACTCGGACCCGACAA 827
              | |   |   |||| ||| |  || |         ||   |||| |||||  ||  ||
Db       6021 CGGCGTACGTGCCCTTCACGCCGAAGGGGCACGTGTCTTTCTCGAGCTCGGGCCTCACGC 6080

Qy        828 CACCCTCACCACCCTCACCCACCACAACCTCCCCAACCCCCCCACCACCACCCTCACCCT 887
              ||||  || ||||  | ||  ||  ||||    |   |     |||  | |   |
Db       6081 TGTCCTCTCCCGCCCTTGCGCAAGACGCCCTCGGACAGGACGAAGGCACGTCGCCATGCGC 6140

Qy        888 CACCCACCCCCACCACCAC 906
              | ||  |||  || || |
Db       6141 CTTCCTTCCCACCCTCCGC 6159


RESULT 12
AAT89956
ID   AAT89956 standard; DNA; 28958 BP.
XX
AC   AAT89956;
XX
DT   12-MAR-1998  (first entry)
XX
DE   Sorangium cellulosum soraphen gene cluster genomic DNA.
XX
KW   Soraphen; polyketide synthase; PKS; antipathogenic; antibiotic;
KW   transgenic plant; phytopathogen; resistance; ss.
XX
OS   Sorangium cellulosum.
XX
PN   US5662898-A.
XX
PD   02-SEP-1997.
XX
PF   01-JUN-1995;   95US-0457342.
XX
PR   20-AUG-1990;   90US-0570184.
PR   02-JUL-1992;   92US-0908284.
PR   31-AUG-1992;   92US-0937648.
PR   01-JUL-1993;   93US-0087636.
PR   08-JUN-1994;   94US-0258261.
XX
PA   (CIBA ) CIBA GEIGY CORP.
XX
PI   Beck JJ,  Gaffney TD,  Hammer PE,  Hill DS,  Lam ST;
PI   Ligon JM,  Ryals JA,  Schupp T,  Uknes SJ;
XX
DR   WPI; 1997-447901/41.
XX
PT   Protecting plants against pathogens with genetically transformed
PT   biological control agent - which expresses all polypeptide(s)
PT   involved in pyrrolnitrin biosynthetic pathway
```

```
XX
PS   Ex 16; Column 104-128; 88pp; English.
XX
CC   This genomic DNA sequence encodes the Soraphen gene cluster. This
CC   sequence encodes two open reading frames (ORF's), the positions of which
CC   are not given in the specification. ORF1 is approximately 25.5 kb in
CC   size and encodes 5 biosynthetic modules homologous to the erythromycin
CC   genes of Saccharopolyspora erythraea. Each module contains a
CC   beta-ketoacylsynthase (KS), an acyltransferase (AT), a ketoreductase
CC   (KR) and an acyl carrier protein (ACP) domain. ORF2 is immediately
CC   adjacent to ORF1 and is thought to contain 3 modules. Soraphen is a
CC   type I polyketide synthase (PKS) which has antibiotic activity.
CC   Transgenic plants containing such antipathogenic genes like those
CC   encoded in the Soraphen cluster should have enhanced resistance to
CC   attack by phytopathogens.
XX
SQ   Sequence 28958 BP; 4084 A; 11863 C; 8356 G; 4655 T; 0 other;

  Query Match            28.8%;  Score 266.2;  DB 18;  Length 28958;
  Best Local Similarity  59.0%;  Pred. No. 5.1e-31;
  Matches  542;  Conservative    0;  Mismatches  353;  Indels   24;  Gaps    4;

Qy          3 CGCATTCATCTGCTCCGGACAGGGCACCCAACGCCCCGGCATGGCCCACGGCCTCTACCA 62
              |||  || ||| | | || || || | ||| || |||  ||||| || ||||||||||| |
Db       5250 CGCCGTCCTCTTCACGGGGCAAGGAAGCCAGCGGCCCACCATGGGCCGCGGCCTCTACGA 5309

Qy         63 CACCCACCCCGTCTTCGCCGCCGCACTCAACGACATCTGCACCCACCTCGACCCCCACCT 122
              |     ||||||||||   | ||| ||| ||  | || || |||||||||||| | | ||
Db       5310 CGTTTTCCCCGTCTTCCGGGACGCCCTCGACACCGTCGGCGCCCACCTCGACCGCGAGCT 5369

Qy        123 CGACCACCCCCTCCTCCCCCTCCTCACCCAAAACGACAACGACAACGAGGACGCGGCCGC 182
              ||||| |||||| | |  | |||||          ||    ||||  |   ||   ||||||
Db       5370 CGACCGCCCCCTGCGCGACGTCCTC------TTCGCTCCCGACGGCTCCGAGCAGGCCGC 5423

Qy        183 ACTGCTCCAGCAGACCCGCTACGCCCAGCCCGCCCTCTTCGCCTTCCAGGTCGCCCTCCA 242
               |  ||| |||| |||  || | ||||||| ||||| || ||| || | |||||||||
Db       5424 GCGCCTCGAGCAAACCGCCTTCACCCAGCCGGCCCTGTTTGCCCTCGAAGTCGCCCTCTT 5483

Qy        243 CCGCCTCCTCACCGACGGCTACCACATCACCCCCCACTACTACGCCGGACACTCCCTCGG 302
              |  || ||     |   || |    | |  |||     |  |  |||||||||| | ||
Db       5484 TCAGCTTCT---ACAATCCTTCGGTCTGAAGCCCGCTCTCCTCCTCGGACACTCCATTGG 5540

Qy        303 CGAAATCACCGCCGCCCACCTCGCCGGCATCCTCACCCTCACCGACGCCACCACCCTCAT 362
              |||  ||  |||||||||| |||||||| |||| | |||   |||| |  |||||| |
Db       5541 CGAGCTCGTCGCCGCCCACGTCGCCGGCGTCCTTTCTCTCCAGGACGGCTGCACCCTCGT 5600

Qy        363 CACCCAACGCGCCACCCTCATGCAAACCATGCCCC---CCGGCACCATGACCACCCTCCA 419
              | ||   ||||| |  |||||||||| |  | || |    |||| |||||  |||||||||
Db       5601 CGCCGCCCGCGCAAAGCTCATGCAAGCGCTCCCACAAGGCGGCGCCATGGTCACCCTCCG 5660

Qy        420 CACCACCCCCCACCACATCACCCACCACCTCACCGCCCACGAAAACGACCTCGCCATCGC 479
              || ||    |  |  ||  | |||  |||    || |||||  |          | ||||
Db       5661 AGCCTCCGAGGAGGAAGTCCGCGACCTTCTCCAGCCCTACGAAGGCCGAGCTAGCCTCGC 5720

Qy        480 CGCCATCAACACCCCCACCTCCCTCGTCATCAGCGGCACCCCCCACACCGTCCAACACAT 539
              ||||  ||||    ||   ||||  |||| ||   |||        || | ||     | ||
Db       5721 CGCCCTCAATGGGCCTCTCTCCACCGTCGTCGCTGGCGATGAAGACGCGGTGGTGGAGAT 5780

Qy        540 CACCACCCTCTGCCAACAACAAGGCATCAAAACCAAAACCCTCCCCACCAACCACGCCTT 599
              | ||  ||    | ||   |  ||    || |||| |  ||| | |  || ||||||||
Db       5781 CGCCCGCCAGGCCGAAGCCCTCGGACGAAAGACCACACGCCTGCGCGTCAGCCACGCCTT 5840

Qy        600 CCACTCCCCCCACACCAACCCCATCCTCAACCAACTCCACCAGCACACCCAAACCCTCAC 659
              ||| ||||| ||||   ||   || ||| || |  ||| ||    | |||| | ||||||
Db       5841 CCATTCCCCGCACATGGACGGAATGCTCGACGACTTCCGCCGCGTCGCCCAGAGCCTCAC 5900

Qy        660 CTACCACCCACCCCACACCCCCCTCATCAC------------CGCCAACACCCCACCCGA 707
              |||||| ||  | | || |||| ||||| |            |||   | || |    | |
Db       5901 CTACCATCCCGCACGCATCCCCATCATCTCCAACGTCACCGGCGCGCGCGCCACGGACCA 5960

Qy        708 CCAACTCCTCACCCCCCACTACTGGACCCAACAAGCCCGCAACACCGTCGACTACGCCAC 767
              | | |||  | | ||| |||||||||  ||  || |  ||| ||||||||  || |  |
Db       5961 CGAGCTCGCCTCGCCCGACTACTGGGTCCGCCACGTTCGCCACACCGTCCGCTTCCTCGA 6020

Qy        768 CACCACCCAAACCCTCCACCAACACGGCGTCACCACCTACATCGAACTCGGACCCGACAA 827
              | |   |  ||||  |||     | || |        ||   |||| ||||| ||  ||
Db       6021 CGGCGTACGTGCCCTTCACGCCGAAGGGGCACGTGTCTTTCTCGAGCTCGGGCCTCACGC 6080

Qy        828 CACCCTCACCACCCTCACCCACCACAACCTCCCCAACCCCCCCACCACCACCCTCACCCT 887
                |||| || ||||  | ||  ||  ||||    |   |     |||  | |    |
Db       6081 TGTCCTCTCCGCCCTTGCGCAAGACGCCCTCGGACAGGACGAAGGCACGTCGCCATGCGC 6140

Qy        888 CACCCACCCCCCACCACCAC 906
              |  ||  |||  || || |
Db       6141 CTTCCTTCCCACCCTCCGC 6159


RESULT 13
AAA75299
ID   AAA75299 standard; DNA; 28958 BP.
XX
AC   AAA75299;
XX
DT   15-JAN-2001  (first entry)
XX
DE   DNA sequence of Sorangium cellulosum soraphen gene region.
XX
KW   Pyrrolnitrin; PrnA; PrnB; PrnC; PrnD; antibiotic; biocontrol;
KW   fungal respiratory electron transport inhibitor; transgenic plant;
KW   antipathogenic substance; biosynthetic gene; fungal resistance;
KW   phytopathogen; soraphen; ss.
XX
OS   Sorangium cellulosum.
XX
PN   US6117670-A.
XX
PD   12-SEP-2000.
XX
PF   24-FEB-1998;   98US-0028934.
XX
PR   30-MAY-1995;   95WO-IB00414.
PR   08-JUN-1994;   94US-0258261.
PR   09-OCT-1996;   96US-0729214.
XX
PA   (NOVS ) NOVARTIS FINANCE CORP.
XX
PI   Ligon JM,  Lam ST,  Hammer PE,  Van Pee K,  Kirner S,  Young TR;
PI   Hill DS;
XX
DR   WPI; 2000-593649/56.
XX
PT   Pyrrolnitrin biosynthesis genes and enzymes, useful for producing
PT   pyrrolnitrin for increasing plant resistance to phytopathogens, e.g.
PT   fungi or nematodes, and for developing microbial hosts useful as
PT   biocontrol agents  -
XX
PS   Example 14; Column 121-146; 124pp; English.
XX
CC   The present sequence represents a Sorangium cellulosum soraphen gene
CC   cluster. The specification describes pyrrolnitrin gene regions. This
CC   region encodes enzymes required in the biosynthetic pathway of
CC   pyrrolnitrin. The PrnA enzyme catalyses the conversion of tryptophan
CC   to 7-chlorotryptophan. The PrnB enzyme catalyses the conversion of
CC   7-chlorotryptophan to monodechloroaminopyrrolnitrin. PrnC enzyme
CC   catalyses the conversion of monodechloroaminopyrrolnitrin to
CC   aminopyrrolnitrin. PrnD enzyme catalyses the conversion of
CC   aminopyrrolnitrin to pyrrolnitrin. Pyrrolnitrin has antibiotic activity,
CC   and inhibits a broad range of fungi. It is a fungal respiratory electron
CC   transport inhibitor. The genes are useful for developing a host as a
CC   biocontrol agent, for protecting plants against a pathogen and for
CC   producing large, uniform amounts of pyrrolnitrin. The genes are also
CC   useful for producing transgenic plants that express antipathogenic
CC   substance (APS) biosynthetic genes. The genes and enzymes are useful
CC   for increasing the resistance of host plants to disease caused by
CC   phytopathogens, e.g. fungi, bacteria or nematodes. The microbial hosts
CC   are useful for controlling or inhibiting the growth of a phytopathogen.
```

```
XX
SQ   Sequence 28958 BP; 4084 A; 11863 C; 8356 G; 4655 T; 0 other;

  Query Match             28.8%;  Score 266.2;  DB 21;  Length 28958;
  Best Local Similarity   59.0%;  Pred. No. 5.1e-31;
  Matches  542;  Conservative    0;  Mismatches  353;  Indels   24;  Gaps    4;

Qy          3 CGCATTCATCTGCTCCGGACAGGGCACCCAACGCCCCGGCATGGCCCACGGCCTCTACCA 62
              |||  || ||| | | || || || | ||| || |||  ||||| || ||||||||||| |
Db       5250 CGCCGTCCTCTTCACGGGGCAAGGAAGCCAGCGGCCCACCATGGGCCGCGGCCTCTACGA 5309

Qy         63 CACCCACCCCGTCTTCGCCGCCGCACTCAACGACATCTGCACCCACCTCGACCCCCACCT 122
              |      ||||||||||    | ||| ||| ||   | || ||| ||||||||||| | | ||
Db       5310 CGTTTTCCCCGTCTTCCGGGACGCCCTCGACACCGTCGGCGCCCACCTCGACCGCGAGCT 5369

Qy        123 CGACCACCCCCTCCTCCCCCTCCTCACCCAAAACGACAACGACAACGAGGACGCGGCCGC 182
              ||||| |||||| | |  | |||||          ||    ||||  |   ||   ||||||
Db       5370 CGACCGCCCCCTGCGCGACGTCCTC------TTCGCTCCCGACGGCTCCGAGCAGGCCGC 5423

Qy        183 ACTGCTCCAGCAGACCCGCTACGCCCAGCCCGCCCTCTTCGCCTTCCAGGTCGCCCTCCA 242
              |  ||| |||| |||   || | ||||||| ||||| || ||| || | |||||||||
Db       5424 GCGCCTCGAGCAAACCGCCTTCACCCAGCCGGCCCTGTTTGCCCTCGAAGTCGCCCTCTT 5483

Qy        243 CCGCCTCCTCACCGACGGCTACCACATCACCCCCCACTACTACGCCGGACACTCCCTCGG 302
              |  || ||      |   || |   | |  |||      |  |  |||||||||| | ||
Db       5484 TCAGCTTCT---ACAATCCTTCGGTCTGAAGCCCGCTCTCCTCCTCGGACACTCCATTGG 5540

Qy        303 CGAAATCACCGCCGCCCACCTCGCCGGCATCCTCACCCTCACCGACGCCACCACCCTCAT 362
              |||  ||  |||||||||| ||||||||  ||||  | |||   |||| |  |||||| |
Db       5541 CGAGCTCGTCGCCGCCCACGTCGCCGGCGTCCTTTCTCTCCAGGACGGCTGCACCCTCGT 5600

Qy        363 CACCCAACGCGCCACCCTCATGCAAACCATGCCCC---CCGGCACCATGACCACCCTCCA 419
              | ||   ||||| |  |||||||||| |   | || |    |||| |||||  ||||||||
Db       5601 CGCCGCCCGCGCAAAGCTCATGCAAGCGCTCCCACAAGGCGGCGCCATGGTCACCCTCCG 5660

Qy        420 CACCACCCCCCACCACATCACCCACCACCTCACCGCCCACGAAAACGACCTCGCCATCGC 479
              || ||    |  |  ||  | |||  |||     || ||||  |         | ||||
Db       5661 AGCCTCCGAGGAGGAAGTCCGCGACCTTCTCCAGCCCTACGAAGGCCGAGCTAGCCTCGC 5720

Qy        480 CGCCATCAACACCCCCACCTCCCTCGTCATCAGCGGCACCCCCCACACCGTCCAACACAT 539
              |||| ||||   ||   ||||  |||| ||   |||       || | ||     | ||
Db       5721 CGCCCTCAATGGGCCTCTCTCCACCGTCGTCGCTGGCGATGAAGACGCGGTGGTGGAGAT 5780

Qy        540 CACCACCCTCTGCCAACAACAAGGCATCAAAACCAAAACCCTCCCCACCAACCACGCCTT 599
              | |||  ||    | ||   |  ||     || |||| |  ||| | |  || ||||||||
Db       5781 CGCCCGCCAGGCCGAAGCCCTCGGACGAAAGACCACACGCCTGCGCGTCAGCCACGCCTT 5840

Qy        600 CCACTCCCCCCACACCAACCCCATCCTCAACCAACTCCACCAGCACACCCAAACCCTCAC 659
              ||| ||||| ||||   ||  || ||| || |  ||| ||    | |||| | ||||||
Db       5841 CCATTCCCCGCACATGGACGGAATGCTCGACGACTTCCGCCCGCGTCGCCCAGAGCCTCAC 5900

Qy        660 CTACCACCCACCCCACACCCCCCTCATCAC------------CGCCAACACCCCACCCGA 707
              |||||| ||  | | ||  |||| ||||| |           |||   | || |    | |
Db       5901 CTACCATCCCGCACGCATCCCCATCATCTCCAACGTCACCGGCGCGCGCGCCACGGACCA 5960

Qy        708 CCAACTCCTCACCCCCCACTACTGGACCCAACAAGCCCGCAACACCGTCGACTACGCCAC 767
              | | |||  | | |||  |||||||||  ||  || |  ||| |||||||||  || | |
Db       5961 CGAGCTCGCCTCGCCCGACTACTGGGTCCGCCACGTTCGCCACACCGTCCGCTTCCTCGA 6020

Qy        768 CACCACCCAAACCCTCCACCAACACGGCGTCACCACCTACATCGAACTCGGACCCGACAA 827
              |  |   |  |||| |||    | || |        ||   |||| ||||| ||  ||
Db       6021 CGGCGTACGTGCCCTTCACGCCGAAGGGGCACGTGTCTTTCTCGAGCTCGGGCCTCACGC 6080

Qy        828 CACCCTCACCACCCTCACCCACCACAACCTCCCCAACCCCCCCACCACCACCCTCACCCT 887
              |||| || ||||  | ||  ||  ||||    |   |     ||| | |     |
Db       6081 TGTCCTCTCCGCCCTTGCGCAAGACGCCCTCGGACAGGACGAAGGCACGTCGCCATGCGC 6140

Qy        888 CACCCACCCCCACCACCAC 906
              |  ||  |||  || || |
Db       6141 CTTCCTTCCCACCCTCCGC 6159


RESULT 14
```

```
AAV05287
ID   AAV05287 standard; DNA; 49377 BP.
XX
AC   AAV05287;
XX
DT   21-MAY-1998  (first entry)
XX
DE   The soraphen biosynthesis gene cluster from Sorangium cellulosum.
XX
KW   Polyketide synthase; PKS; biosynthesis; soraphen; SorR; SorA; SorB;
KW   SorM; biosynthetic module; beta-ketoacylsynthase; acyltransferase;
KW   ketoreductase; beta-ketone processing domain; cytostatic agent;
KW   antimicrobial agent; phytopathogenic fungi; transgenic plant;
KW   biological control; ss.
XX
OS   Sorangium cellulosum.
XX
FH   Key             Location/Qualifiers
FT   CDS             383..760
FT                   /*tag= a
FT                   /product= SorR
FT                   /note= "gene product highly homologous to the
FT                          reductase domains of type I PKSs such
FT                          as eryA from Saccharopolyspora erythraea"
FT   CDS             927..19874
FT                   /*tag= b
FT                   /product= SorA
FT                   /note= "gene product is highly homologous to
FT                          type I PKSs that are known to be involved
FT                          in the synthesis of polyketide compounds"
FT   misc_feature    942..7115
FT                   /*tag= c
FT                   /note= "module 1 of SorA"
FT   misc_feature    7203..12884
FT                   /*tag= d
FT                   /note= "module 2 of SorA"
FT   misc_feature    13455..19616
FT                   /*tag= e
FT                   /note= "module 3 of SorA"
FT   misc_feature    19871..46318
FT                   /*tag= f
FT   CDS             19871..46318
FT                   /*tag= g
FT                   /product= SorB
FT                   /note= "gene product is highly homologous to
FT                          type I PKS genes"
FT   misc_feature    19870..24556
FT                   /*tag= h
FT                   /note= "module 1 of SorB"
FT   misc_feature    24638..30820
FT                   /*tag= i
FT                   /note= "module 2 of SorB"
FT   misc_feature    30881..35446
FT                   /*tag= j
FT                   /note= "module 3 of SorB"
FT   misc_feature    35528..40114
FT                   /*tag= k
FT                   /note= "module 4 of SorB"
FT   misc_feature    40190..46318
FT                   /*tag= l
FT                   /note= "module 5 of SorB"
FT   CDS             46851..47891
FT                   /*tag= m
FT                   /product= SorM
FT                   /note= "gene product is homologous to the
FT                          methyltransferase from Streptomyces
FT                          hygroscopicus that is involved in
FT                          the synthesis of the polyketide rappamicin"
XX
PN   US5716849-A.
XX
PD   10-FEB-1998.
XX
```

```
PF   14-DEC-1996;   96US-0764233.
XX
PR   24-AUG-1993;   93WO-US07954.
PR   08-JUN-1994;   94US-0258261.
PR   09-OCT-1996;   96US-0729214.
XX
PA   (NOVS ) NOVARTIS FINANCE CORP.
XX
PI   Beck JJ, Hill DS, Ligon JM, Neff S, Ryals JA, Schupp T;
XX
DR   WPI; 1998-158369/14.
XX
PT   DNA encoding Sorangium cellulosum polypeptide(s) - used for, e.g.
PT   biosynthesis of soraphen useful as antimicrobial agent against
PT   phytopathogenic fungi
XX
PS   Claim 2; Columns 47-90; 64pp; English.
XX
CC   The present sequence contains a cluster of genes that encode polyketide
CC   synthases (PKSs) that are involved in the synthesis of soraphens in
CC   Sorangium cellulosum. The proteins encoded by the present sequence are
CC   SorR, SorA, SorB and SorM. SorA and SorB contain biosynthetic modules
CC   which contain a beta-ketoacylsynthase, an acyltransferase, a
CC   ketoreductase and an acyl carrier protein domain, as well as beta-ketone
CC   processing domains. S. cellulosum soraphens are useful as a cytostatic
CC   and antimicrobial agent active against phytopathogenic fungi.
CC   Soraphen-producing transgenic plants or biological control agents can
CC   also be produced, which may reduce crop losses and nutritional
CC   deprivation for local populations in many parts of the world.
XX
SQ   Sequence 49377 BP; 7247 A; 19522 C; 14477 G; 8131 T; 0 other;

  Query Match             28.8%;  Score 266.2;  DB 19;  Length 49377;
  Best Local Similarity   59.0%;  Pred. No. 4.8e-31;
  Matches  542;  Conservative    0;  Mismatches  353;  Indels   24;  Gaps    4;

Qy          3 CGCATTCATCTGCTCCGGACAGGGCACCCAACGCCCCGGCATGGCCCACGGCCTCTACCA 62
              ||| || ||| | | || || || | ||| || |||  ||||| || || ||||||| |
Db       8801 CGCCGTCCTCTTCACGGGGCAAGGAAGCCAGCGGCCCACCATGGGCCGCGCCCTCTACGA 8860

Qy         63 CACCCACCCCGTCTTCGCCGCCGCACTCAACGACATCTGCACCCACCTCGACCCCCACCT 122
              | |   ||||||||||   | ||| ||| ||  | ||  | ||||||||||||| | ||||
Db       8861 CGCTTTCCCCGTCTTCCGGGACGCCCTCGACACCGTCGCCGCCCACCTCGACCGCGACCT 8920

Qy        123 CGACCACCCCCTCCTCCCCCTCCTCACCCAAAACGACAACGACAACGAGGACGCGGCCGC 182
              ||||| ||||||  | |  | |||||         ||    ||||  |   ||  ||||||
Db       8921 CGACCGCCCCCTGCGCGACGTCCTC------TTCGCTCCCGACGGCTCCGAGCAGGCCGC 8974

Qy        183 ACTGCTCCAGCAGACCCGCTACGCCCAGCCCGCCCTCTTCGCCTTCCAGGTCGCCCTCCA 242
              |  ||| | || |||  || | ||||||| ||||| || ||| || | |||||||||
Db       8975 GCGCCTCGACCAAACCGCCTTCACCCAGCCGGCCCTGTTTGCCCTCGAAGTCGCCCTCTT 9034

Qy        243 CCGCCTCCTCACCGACGGCTACCACATCACCCCCCACTACTACGCCGGACACTCCCTCGG 302
              |  || ||   | |   || |    | |  |||     |  |  |||||||||| | ||
Db       9035 TCAGCTTCT---CCAATCCTTCGGTCTGAAGCCCGCTCTCCTCCTCGGACACTCCATTGG 9091

Qy        303 CGAAATCACCGCCGCCCACCTCGCCGGCATCCTCACCCTCACCGACGCCACCACCCTCAT 362
              |||  ||  |||||||||| |||||||| ||||  | |||   ||||||  ||||||| |
Db       9092 CGAGCTCGTCGCCGCCCACGTCGCCGGCGTCCTTTCTCTCCAGGACGCCTGCACCCTCGT 9151

Qy        363 CACCCAACGCGCCACCCTCATGCAAACCATGCCCC---CCGGCACCATGACCACCCTCCA 419
              | ||   ||||| |  |||||||| |  | || |   |||| |||||  ||||||||
Db       9152 CGCCGCCCGCGCAAAGCTCATGCAAGCGCTCCCACAAGGCGGCGCCATGGTCACCCTCCG 9211

Qy        420 CACCACCCCCCACCACATCACCCACCACCTCACCGCCCACGAAAACGACCTCGCCATCGC 479
              || ||    |  |  ||  | |||  |||    || ||||                | ||||
Db       9212 AGCCTCCGAGGAGGAAGTCCGCGACCTTCTCCAGCCCTACGATGGACGAGCTAGCCTCGC 9271

Qy        480 CGCCATCAACACCCCCACCTCCCTCGTCATCAGCGGCACCCCCCACACCGTCCAACACAT 539
              |||| ||||    ||   ||||  |||| ||   |||       || | ||     | ||
Db       9272 CGCCCTCAATGGGCCTCTCTCCACCGTCGTCGCTGGCGATGAAGACGCGGTGGTGGAGAT 9331

Qy        540 CACCACCCTCTGCCAACAACAAGGCATCAAAACCAAAACCCTCCCCACCAACCACGCCTT 599
              | ||  ||    | ||   |  ||    || |||| |  ||| | |  || |||||||||
Db       9332 CGCCCGCCAGGCCGAAGCCCTCGGACGAAAGACCACACGCCTGCGCGTCAGCCACGCCTT 9391

Qy        600 CCACTCCCCCCACACCAACCCCATCCTCAACCAACTCCACCAGCACACCCAAACCCTCAC 659
              |||||| || ||||    ||   || ||| || |  ||| ||     | |||| | ||||||
Db       9392 CCACTCTCCGCACATGGACGGAATGCTCGACGACTTCCGCCGCGTCGCCCAGAGCCTCAC 9451

Qy        660 CTACCACCCACCCCACACCCCCCTCATCAC------------CGCCAACACCCCACCCGA 707
              |||||| ||  | | || |||| ||||| |             |||   | ||  |  | |
Db       9452 CTACCATCCCGCACGCATCCCCATCATCTCCAACGTCACCGGCGCGCGCGCCACGGACCA 9511

Qy        708 CCAACTCCTCACCCCCCACTACTGGACCCAACAAGCCCGCAACACCGTCGACTACGCCAC 767
              | | |||  | | ||| ||||||||  ||  || |  ||| ||||||||  || |  |
Db       9512 CGAGCTCGCCTCGCCCGACTACTGGGTCCGCCACGTTCGCCACACCGTCCGCTTCCTCGA 9571

Qy        768 CACCACCCAAACCCTCCACCAACACGGCGTCACCACCTACATCGAACTCGGACCCGACAA 827
              | |   |   |||| |||    | || |    |  ||   |||| ||||| ||  ||
Db       9572 CGGCGTACGTGCCCTTCACGCCGAAGGGGCACGCGTCTTTCTCGAGCTCGGGCCTCACGC 9631

Qy        828 CACCCTCACCACCCTCACCCACCACAACCTCCCCAACCCCCCCCACCACCACCCTCACCCT 887
              ||||  || ||||  | ||  ||  ||||    |   |     |||  | |   |  .
Db       9632 TGTCCTCTCCGCCCTTGCGCAAGACGCCCTCGGACAGGACGAAGGCACGTCGCCATGCGC 9691

Qy        888 CACCCACCCCCCACCACCAC 906
              |  ||  |||  || || |
Db       9692 CTTCCTTCCCACCCTCCGC 9710


RESULT 15
AAV99229
ID   AAV99229 standard; DNA; 925 BP.
XX
AC   AAV99229;
XX
DT   09-MAR-1999  (first entry)
XX
DE   DNA encoding an active acyltransferase domain (LigAT) from a PKS.
XX
KW   Enzymatically active acyltransferase domain; polyketide synthase; PKS;
KW   polyketide synthesis; antimicrobial; anticancer; antifungal;
KW   immunosuppressant; anthelmintic; LigAT; malonyl acyltransferase domain;
KW   module 2; ligase-PKS cluster; Streptomyces hygroscopicus ATCC29253; ds.
XX
OS   Streptomyces hygroscopicus.
XX
PN   WO9851695-A2.
XX
PD   19-NOV-1998.
XX
PF   13-MAY-1998;   98WO-US09518.
XX
PR   16-MAY-1997;   97US-0858003.
XX
PA   (ABBO ) ABBOTT LAB.
XX
PI   Kakavas SJ,  Katz L,  Pereda-Lopez A,  Ruan X,  Stassi DL;
PI   Summers RG;
XX
DR   WPI; 1999-070114/06.
DR   P-PSDB; AAW87712.
XX
PT   New polyketides produced by microorganism having altered polyketide
PT   synthase genes - especially chimeras containing a heterologous
PT   acyltransferase domain, useful as e.g. antimicrobials, anticancer
PT   agents
XX
PS   Claim 24; Fig 7; 137pp; English.
XX
CC   The present sequence encodes an enzymatically active acyltransferase
CC   domain, LigAT, from a polyketide synthase (PKS). The LigAT protein
CC   is derived from the malonyl acyltransferase domain from module 2 of the
CC   ligase-PKS cluster of Streptomyces hygroscopicus ATCC29253. The sequence
CC   can be used to replace acyltransferase domains of other PKS enzymes,
```

CC which are then used to generate novel polyketides by de novo biosynthesis
CC rather than by chemical modification. Polyketides of the invention have
CC the methyl groups on the macrolactone ring substituted with -H, -Et
CC and/or -OH, and the ethyl side chain substituted with hydroxymethyl or
CC dihydroxycyclohexylmethyl side chains. Polyketides of the invention are
CC antimicrobial, anticancer, antifungal, immunosuppressant and/or
CC anthelmintic agents, or their intermediates.
XX
SQ Sequence 925 BP; 188 A; 448 C; 179 G; 110 T; 0 other;

Query Match 27.4%; Score 253; DB 20; Length 925;
Best Local Similarity 61.0%; Pred. No. 5.8e-29;
Matches 558; Conservative 0; Mismatches 255; Indels 102; Gaps 5;

```
Qy          4 GCATTCATCTGCTCCGGACAGGGCACCCAACGCCCCGGCATGGCCCACGGCCTCTACCAC 63
              ||  | || | | |||||||||||| |||||||||||||||||  |      | ||| |
Db         10 GCGGTGATGTTCACCGGACAGGGCTCCCAACGCCCCGGCATGGGACGACAGTTGTACGAG 69

Qy         64 ACCCACCCCGTCTTCGCCGCCGCACTCAACGACATCTGCACCCACCTCGACCCCCACCTC 123
                | ||||||||||||   |||||  ||||  ||| | | | |  |||| |  ||
Db         70 CACTTCCCCGTCTTCGCCCAGGCACTGGACGAGGTCTTCGCACTCGCCACCCCCGGACTA 129

Qy        124 GACCACCCCCTCCTCCCCCTCCTCACCCAAAACGACAACGACAACGAGGACGCGGCCGCA 183
              | |     |  ||  || |                            |||  ||||| |
Db        130 CGCGAGGTGATGTTCGACCCC--------------------------GACCAGGCCGAA 162

Qy        184 CTGCTCCAGCAGACCCGCTACGCCCAGCCCGCCCTCTTCGCCTTCCAGGTCGCCCTCCAC 243
              |||||  |  |||  | ||||||||  ||||||  |||||||||  |   ||||||| ||
Db        163 ACACTCCAACGCACCGACCACGCCCAGATCGCCCTGTTCGCCTTCGAAACCGCCCTCTAC 222

Qy        244 CGCCTCCTCACCGACGGCTACCACATCACCCCCCACTACTACGCCGGACACTCCCTCGGC 303
              ||    |||   ||   ||    | |     |||  ||      |  |||||||||  ||||
Db        223 CG---ACTCTGGGAATCCTGGGGCCTGCGACCCGACATGGTCTGCGGACACTCGGTCGGA 279

Qy        304 GAAATCACCGCCGCCCACCTCGCCGGCATCCTCACCCTCACCGACGCCACCACCCTCATC 363
              |||||||||||  ||||||  ||  ||||||  ||||||||||  ||||||||   |  |||| ||
Db        280 GAAATCACCGCAGCCCACGTCTCCGGCACCCTCACCCTCCCCGACGCCGTCCACCTCGTC 339

Qy        364 ACCCAACGCGCCACCCTCATGCAAACCATGCCCCC---CGGCACCATGACCACCCTCCAC 420
              |||   |||||  ||||||||||||||  |  |||||||    ||||  |||||   |  ||  ||   |
Db        340 ACCACACGCGGCACCCTCATGCAAAACCTGCCCCCCCGGCGGCGCCATGCTCGCCGTCGCC 399

Qy        421 ACCACCCCCCACCACATCACCCACCACCTCACCGCCCACGAAAACGACCTCGCCATCGCC 480
              |||  |||||||   |  ||   | |||||||   |   ||||  |   ||   |  ||  ||||||||
Db        400 ACCGACCCCCACACCCTCCAACCCCACCTCGACAACCACCACGACACCATCTCCATCGCC 459

Qy        481 GCCATCAACACCCCCACCTCCCTCGTCATCAGCGGCACCCCCCACACCGTCCAACACATC 540
              |||||||||  ||||  |  ||  ||||  ||  ||||  ||  |  ||||  ||||  ||||||
Db        460 GCCATCAACGGCCCCCACGCCACCGTCCTCTCCGGCGACCGCACCACCCTCCACCACATC 519

Qy        541 ACCACCCTCTGCCAACAACAAGGCATCAAAACCAAAACCCTCCCCACCAACCACGCCTTC 600
              ||||||                  ||||  ||||||  ||  ||  ||||||  ||  |  |
Db        520 GCCACCCAAC--------------TCAACACCAAACCCTTCACCACCACCCTCAAC--- 561

Qy        601 CACTCCCCCCACACCAACCCCATCCTCAACCAACTCCACCAGCACACCCAAACCCTCACC 660
                                                                  |||||||||
Db        562 ----------------------------------------------------ACCCTCACC 570

Qy        661 TACCACCCACCCCACACCCCCCTCATCACCGCCAACACCCCACCCGACCAACTCCTCACC 720
              |||||||  ||  |||||  ||||||||||  |      ||||  |   |    |   | ||  |  |
Db        571 CACCACCCCCCACACACACCCCTCATCAGCATGCTCACCGCCACACCCACCCACCCCGAC 630

Qy        721 CCCCACTACTGGACCCAACAAGCCCGCAACACCGTCGACTACGCCACCACCACCCAAACC 780
              ||  |  ||||||||||  ||    |   |     |||||  ||||  ||   ||||   |||    |
Db        631 ACCACCCACTGGACCCAGCACATCACCGCACCCGTCCGCTACACCGACACCCTCCACCAC 690

Qy        781 CTCCACCAACACGGCGTCACCACCTACATCGAACTCGGACCCGACAACACCCTCACCACC 840
              ||||||||  ||||||  ||||||||||||  |||||  ||||  |||||||  ||||||||||  ||
Db        691 CTCCACCACCACGGCATCACCACCTACCTCGAAATCGGCCCCGACACCACCCTCACCGCC 750

Qy        841 CTCACCCACCACAACCTCCCCAACCCCCCCACCACCACCCTCACCCTCACCCACCCCCAC 900
              |||  |||  |   ||  ||||||||  |  ||  ||   |  ||  ||  ||||   |  ||  |  ||||
Db        751 CTCGCCCGCACCACCCTCCCCACCACCACCCACCTCATCCCCACCACCCGCCGCAACCAC 810
```

```
Qy        901 CACCACCCCCAAACC 915
                  || |   ||  | |
Db        811 AACGAAGTCCGCAGC 825
```

Search completed: June 17, 2003, 13:48:41
Job time : 189.899 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 17, 2003, 11:45:13 ; Search time 40.0267 Seconds
(without alignments)
7079.508 Million cell updates/sec

Title: US-09-914-286-1_COPY_7906_8829
Perfect score: 924
Sequence: 1 accgcattcatctgctccgg.....acccccaaacccacctcctc 924

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	266.2	28.8	28958	1	US-08-258-261B-6	Sequence 6, Appli
	2	266.2	28.8	28958	1	US-08-456-837-6	Sequence 6, Appli
	3	266.2	28.8	28958	1	US-08-457-342-6	Sequence 6, Appli
	4	266.2	28.8	28958	1	US-08-457-646A-6	Sequence 6, Appli
	5	266.2	28.8	28958	1	US-08-458-076A-6	Sequence 6, Appli
	6	266.2	28.8	28958	1	US-08-764-233A-4	Sequence 4, Appli
	7	266.2	28.8	28958	1	US-08-457-335A-6	Sequence 6, Appli
	8	266.2	28.8	28958	1	US-08-729-214-6	Sequence 6, Appli
	9	266.2	28.8	28958	3	US-09-028-934-6	Sequence 6, Appli
	10	266.2	28.8	49377	1	US-08-764-233A-1	Sequence 1, Appli
	11	253	27.4	925	3	US-08-858-003-1	Sequence 1, Appli
	12	253	27.4	925	3	US-09-078-166-1	Sequence 1, Appli
	13	253	27.4	925	4	US-08-997-467-1	Sequence 1, Appli
	14	239.4	25.9	15872	4	US-09-105-537-1	Sequence 1, Appli
	15	232.6	25.2	50937	4	US-09-428-517-1	Sequence 1, Appli
	16	220.2	23.8	13842	4	US-09-105-537-30	Sequence 30, Appl
	17	220.2	23.8	36778	4	US-09-105-537-5	Sequence 5, Appli
	18	220.2	23.8	38506	3	US-09-320-878-19	Sequence 19, Appl
	19	219.4	23.7	43280	2	US-08-804-227C-1	Sequence 1, Appli
	20	218.8	23.7	1030	3	US-08-858-003-2	Sequence 2, Appli
	21	218.8	23.7	1030	3	US-09-078-166-2	Sequence 2, Appli
	22	218.8	23.7	1030	4	US-08-997-467-2	Sequence 2, Appli
	23	198.8	21.5	44377	2	US-08-804-227C-7	Sequence 7, Appli
	24	198.8	21.5	44377	2	US-08-804-198-1	Sequence 1, Appli
	25	192.2	20.8	4041	4	US-09-105-537-36	Sequence 36, Appl
	26	179.8	19.5	1926	4	US-09-249-585A-4	Sequence 4, Appli
	27	179.8	19.5	1931	2	US-09-130-114-2	Sequence 2, Appli
	28	176.6	19.1	4689	4	US-09-105-537-34	Sequence 34, Appl
c	29	164.4	17.8	1881	4	US-09-434-288-5	Sequence 5, Appli
	30	162.2	17.6	11220	4	US-09-105-537-32	Sequence 32, Appl
c	31	157	17.0	1590	4	US-09-434-288-1	Sequence 1, Appli
c	32	146	15.8	13987	2	US-08-804-227C-13	Sequence 13, Appl
c	33	146	15.8	44377	2	US-08-804-227C-7	Sequence 7, Appli
c	34	146	15.8	44377	2	US-08-804-198-1	Sequence 1, Appli
	35	134.2	14.5	5163	3	US-08-700-651-1	Sequence 1, Appli
	36	134.2	14.5	5163	3	US-08-928-361B-4	Sequence 4, Appli
	37	134.2	14.5	5318	3	US-08-700-651-2	Sequence 2, Appli
	38	134.2	14.5	5318	3	US-08-928-361B-3	Sequence 3, Appli
c	39	130.4	14.1	1505	1	US-07-915-246-1	Sequence 1, Appli
c	40	130.2	14.1	4403765	4	US-09-103-840A-2	Sequence 2, Appli
	41	129.4	14.0	11219	1	US-07-642-734C-1	Sequence 1, Appli
	42	129.4	14.0	11219	3	US-08-439-009A-1	Sequence 1, Appli
	43	127.2	13.8	71989	4	US-09-443-501A-2	Sequence 2, Appli
c	44	125.8	13.6	4411529	4	US-09-103-840A-1	Sequence 1, Appli
	45	124.6	13.5	5511	3	US-08-928-361B-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1
US-08-258-261B-6
; Sequence 6, Application US/08258261B
; Patent No. 5639949
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James M.
; APPLICANT: Beck, James Joseph
; APPLICANT: Hill, Dwight Steven
; APPLICANT: Ryals, John Andrew
; APPLICANT: Gaffney, Thomas Deane
; APPLICANT: Lam, Stephen Ting
; APPLICANT: Hammer, Phillip E.
; APPLICANT: Uknes, Scott Joseph
; TITLE OF INVENTION: Genes for the synthesis of
; TITLE OF INVENTION: antipathogenic substances
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ciba-Geigy Corporation
; STREET: 7 Skyline Drive
; CITY: Hawthorne
; STATE: NY
; COUNTRY: USA
; ZIP: 10532
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/258,261B
; FILING DATE: 08-JUN-1994
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/457,205
; FILING DATE: 01-JUN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Elmer, James Scott
; REGISTRATION NUMBER: 36,129
; REFERENCE/DOCKET NUMBER: CGC 1506/CIP3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8614
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 28958 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)

```
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
US-08-258-261B-6

  Query Match             28.8%;  Score 266.2;  DB 1;  Length 28958;
  Best Local Similarity   59.0%;  Pred. No. 6.4e-37;
  Matches  542;  Conservative    0;  Mismatches  353;  Indels   24;  Gaps    4;

Qy          3 CGCATTCATCTGCTCCGGACAGGGCACCCAACGCCCCGGCATGGCCCACGGCCTCTACCA 62
              ||| || ||| | | || || || | ||| || |||  ||||| || ||||||||||| |
Db       5250 CGCCGTCCTCTTCACGGGGCAAGGAAGCCAGCGGCCCACCATGGGCCGCGGCCTCTACGA 5309

Qy         63 CACCCACCCCGTCTTCGCCGCCGCACTCAACGACATCTGCACCCACCTCGACCCCCACCT 122
              |     ||||||||||    | ||| ||| ||  | || || ||||||||||||| | | ||
Db       5310 CGTTTTCCCCGTCTTCCGGGACGCCCTCGACACCGTCGGCGCCCACCTCGACCGCGAGCT 5369

Qy        123 CGACCACCCCCCTCCTCCCCCTCCTCACCCAAAACGACAACGACAACGAGGACGCGGCCGC 182
              ||||| ||||||  | |  | |||||         ||    |||| |   ||   ||||||
Db       5370 CGACCGCCCCCTGCGCGACGTCCTC------TTCGCTCCCGACGGCTCCGAGCAGGCCGC 5423

Qy        183 ACTGCTCCAGCAGACCCGCTACGCCCAGCCCGCCCTCTTCGCCTTCCAGGTCGCCCTCCA 242
              |  ||| |||| |||  || | ||||||| ||||| || ||| || | |||||||||
Db       5424 GCGCCTCGAGCAAACCGCCTTCACCCAGCCGGCCCTGTTTGCCCTCGAAGTCGCCCTCTT 5483

Qy        243 CCGCCTCCTCACCGACGGCTACCACATCACCCCCCACTACTACGCCGGACACTCCCTCGG 302
              |  || ||      |   || |    | |  |||      |  |  |||||||||| | ||
Db       5484 TCAGCTTCT---ACAATCCTTCGGTCTGAAGCCCGCTCTCCTCCTCGGACACTCCATTGG 5540

Qy        303 CGAAATCACCGCCGCCCACCTCGCCGGCATCCTCACCCTCACCGACGCCACCACCCTCAT 362
              |||  ||  |||||||||| |||||||| ||||  | |||   |||| |  ||||||| |
Db       5541 CGAGCTCGTCGCCGCCCACGTCGCCGGCGTCCTTTCTCTCCAGGACGGCTGCACCCTCGT 5600

Qy        363 CACCCAACGCGCCACCCTCATGCAAACCATGCCCC---CCGGCACCATGACCACCCTCCA 419
              | ||   ||||| |  |||||||||| |  | || |    |||| |||||  |||||||||
Db       5601 CGCCGCCCGCGCAAAGCTCATGCAAGCGCTCCCACAAGGCGGCGCCATGGTCACCCTCCG 5660

Qy        420 CACCACCCCCCACCACATCACCCACCACCTCACCGCCCACGAAAACGACCTCGCCATCGC 479
              || ||    |  | ||  | ||| |||    || |||||  |         | ||||
Db       5661 AGCCTCCGAGGAGGAAGTCCGCGACCTTCTCCAGCCCTACGAAGGCCGAGCTAGCCTCGC 5720

Qy        480 CGCCATCAACACCCCCACCTCCCTCGTCATCAGCGGCACCCCCCCACACCGTCCAACACAT 539
              |||| ||||     ||   ||||  |||| ||   |||        || | ||     | ||
Db       5721 CGCCCTCAATGGGCCTCTCTCCACCGTCGTCGCTGGCGATGAAGACGCGGTGGTGGAGAT 5780

Qy        540 CACCACCCTCTGCCAACAACAAGGCATCAAAACCAAAACCCTCCCCACCAACCACGCCTT 599
              | ||  ||    | ||    |  ||     || |||| |  ||| | |  || |||||||||
Db       5781 CGCCCGCCAGGCCGAAGCCCTCGGACGAAAGACCACACGCCTGCGCGTCAGCCACGCCTT 5840

Qy        600 CCACTCCCCCCACACCAACCCCATCCTCAACCAACTCCACCAGCACACCCAAACCCTCAC 659
              ||| ||||| ||||    ||   || ||| || |  ||| ||     | |||| | ||||||
Db       5841 CCATTCCCCGCACATGGACGGAATGCTCGACGACTTCCGCCGCGTCGCCCAGAGCCTCAC 5900

Qy        660 CTACCACCCACCCCACACCCCCCTCATCAC------------CGCCAACACCCCACCCGA 707
              |||||| ||  |  | || |||| ||||| |            |||   | || |   | |
Db       5901 CTACCATCCCGCACGCATCCCCATCATCTCCAACGTCACCGGCGCGCGCGCCACGGACCA 5960

Qy        708 CCAACTCCTCACCCCCCACTACTGGACCCAACAAGCCCGCAACACCGTCGACTACGCCAC 767
              | | |||  | | ||| ||||||||  ||  || |  ||| ||||||||  || |  |
Db       5961 CGAGCTCGCCTCGCCCGACTACTGGGTCCGCCACGTTCGCCACACCGTCCGCTTCCTCGA 6020

Qy        768 CACCACCCAAACCCTCCACCAACACGGCGTCACCACCTACATCGAACTCGGACCCGACAA 827
              |  |   |  |||| |||     | ||  |         ||  |||| ||||| ||  ||
Db       6021 CGGCGTACGTGCCCTTCACGCCGAAGGGGCACGTGTCTTTCTCGAGCTCGGGCCTCACGC 6080

Qy        828 CACCCTCACCACCCTCACCCACCACAACCTCCCCAACCCCCCCACCACCACCCTCACCCT 887
              |||| || ||||  | ||  ||  ||||     |   |     |||  | |    |
Db       6081 TGTCCTCTCCGCCCTTGCGCAAGACGCCCTCGGACAGGACGAAGGCACGTCGCCATGCGC 6140

Qy        888 CACCCACCCCCCACCACCAC 906
              |  ||  |||  || || |
Db       6141 CTTCCTTCCCACCCTCCGC 6159
```

```
RESULT 2
US-08-456-837-6
; Sequence 6, Application US/08456837
; Patent No. 5643774
;  GENERAL INFORMATION:
;    APPLICANT:  Schupp, Thomas
;    APPLICANT:  Ligon, James M.
;    APPLICANT:  Beck, James Joseph
;    APPLICANT:  Hill, Dwight Steven
;    APPLICANT:  Ryals, John Andrew
;    APPLICANT:  Gaffney, Thomas Deane
;    APPLICANT:  Lam, Stephen Ting
;    APPLICANT:  Hammer, Phillip E.
;    APPLICANT:  Uknes, Scott Joseph
;    TITLE OF INVENTION:  Genes for the synthesis of
;    TITLE OF INVENTION:  antipathogenic substances
;    NUMBER OF SEQUENCES:  22
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Ciba-Geigy Corporation
;      STREET:  7 Skyline Drive
;      CITY:  Hawthorne
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10532
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/456,837
;      FILING DATE:  01-JUN-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/457,205
;      FILING DATE:  01-JUN-1995
;      APPLICATION NUMBER:  08/258,261
;      FILING DATE:  08-Jun-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Elmer, James Scott
;      REGISTRATION NUMBER:  36,129
;      REFERENCE/DOCKET NUMBER:  CGC 1506/CIP3
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-541-8614
;      TELEFAX:  919-541-8689
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  28958 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
US-08-456-837-6

  Query Match             28.8%;  Score 266.2;  DB 1;  Length 28958;
  Best Local Similarity   59.0%;  Pred. No. 6.4e-37;
  Matches  542;  Conservative    0;  Mismatches  353;  Indels   24;  Gaps    4;

Qy          3 CGCATTCATCTGCTCCGGACAGGGCACCCAACGCCCCGGCATGGCCCACGGCCTCTACCA 62
              ||| || ||| | | || || || | ||| || |||  ||||| || ||||||||||| |
Db       5250 CGCCGTCCTCTTCACGGGGCAAGGAAGCCAGCGGCCCACCATGGGCCGCGGCCTCTACGA 5309

Qy         63 CACCCACCCCGTCTTCGCCGCCGCACTCAACGACATCTGCACCCACCTCGACCCCCACCT 122
              |     ||||||||||    | ||| ||| ||  | || || ||||||||||||| | | ||
Db       5310 CGTTTTCCCCGTCTTCCGGGACGCCCTCGACACCGTCGGCGCCCACCTCGACCGCGAGCT 5369

Qy        123 CGACCACCCCCCTCCTCCCCCTCCTCACCCAAAACGACAACGACAACGAGGACGCGGCCGC 182
              ||||| ||||||  | |  | |||||         ||    |||| |   ||   ||||||
Db       5370 CGACCGCCCCCTGCGCGACGTCCTC------TTCGCTCCCGACGGCTCCGAGCAGGCCGC 5423

Qy        183 ACTGCTCCAGCAGACCCGCTACGCCCAGCCCGCCCTCTTCGCCTTCCAGGTCGCCCTCCA 242
```

```
Db       5424 GCGCCTCGAGCAAACCGCCTTCACCCAGCCGGCCCTGTTTGCCCTCGAAGTCGCCCTCTT 5483

Qy        243 CCGCCTCCTCACCGACGGCTACCACATCACCCCCCACTACTACGCCGGACACTCCCTCGG 302
              |  || ||     |   || |    | |  |||      |  |  |||||||||| | ||
Db       5484 TCAGCTTCT---ACAATCCTTCGGTCTGAAGCCCGCTCTCCTCCTCGGACACTCCATTGG 5540

Qy        303 CGAAATCACCGCCGCCCACCTCGCCGGCATCCTCACCCTCACCGACGCCACCACCCTCAT 362
              |||  ||  |||||||||| |||||||| ||||  | |||    |||| |  |||||||
Db       5541 CGAGCTCGTCGCCGCCCACGTCGCCGGCGTCCTTTCTCTCCAGGACGGCTGCACCCTCGT 5600

Qy        363 CACCCAACGCGCCACCCTCATGCAAACCATGCCCC---CCGGCACCATGACCACCCTCCA 419
              | ||   |||||  |  |||||||||  |  | || |    |||| |||||  ||||||||
Db       5601 CGCCGCCCGCGCAAAGCTCATGCAAGCGCTCCCACAAGGCGGCGCCATGGTCACCCTCCG 5660

Qy        420 CACCACCCCCCACCACATCACCCACCACCTCACCGCCCACGAAAACGACCTCGCCATCGC 479
              ||  ||    |  |  ||  | |||  |||     || |||||  |        | ||||
Db       5661 AGCCTCCGAGGAGGAAGTCCGCGACCTTCTCCAGCCCTACGAAGGCCGAGCTAGCCTCGC 5720

Qy        480 CGCCATCAACACCCCCACCTCCCTCGTCATCAGCGGCACCCCCCACACCGTCCAACACAT 539
              ||||  ||||     ||   ||||  |||| ||   |||       || | ||      | ||
Db       5721 CGCCCTCAATGGGCCTCTCTCCACCGTCGTCGCTGGCGATGAAGACGCGGTGGTGGAGAT 5780

Qy        540 CACCACCCTCTGCCAACAACAAGGCATCAAAACCAAAACCCTCCCCACCAACCACGCCTT 599
              | ||  ||    | ||    |  ||     || |||| |  |||  | |  || |||||||||
Db       5781 CGCCCGCCAGGCCGAAGCCCTCGGACGAAAGACCACACGCCTGCGCGTCAGCCACGCCTT 5840

Qy        600 CCACTCCCCCCACACCAACCCCATCCTCAACCAACTCCACCAGCACACCCAAACCCTCAC 659
              |||  ||||| ||||   ||   || ||| || |  |||  ||    | |||| | ||||||
Db       5841 CCATTCCCCGCACATGGACGGAATGCTCGACGACTTCCGCCGCGTCGCCCAGAGCCTCAC 5900

Qy        660 CTACCACCCACCCCACACCCCCCTCATCAC------------CGCCAACACCCCACCCGA 707
              ||||||  ||   |  |  || |||| |||||  |             |||    | ||  |   | |
Db       5901 CTACCATCCCGCACGCATCCCCATCATCTCCAACGTCACCGGCGCGCGCGCCACGGACCA 5960

Qy        708 CCAACTCCTCACCCCCCACTACTGGACCCAACAAGCCCGCAACACCGTCGACTACGCCAC 767
              | | |||  | | |||  ||||||||  ||  || |  ||| ||||||||  || | |
Db       5961 CGAGCTCGCCTCGCCCGACTACTGGGTCCGCCACGTTCGCCACACCGTCCGCTTCCTCGA 6020

Qy        768 CACCACCCAAACCCTCCACCAACACGGCGTCACCACCTACATCGAACTCGGACCCGACAA 827
              |  |   |   |||| |||    | || |       ||   |||| ||||| ||  ||
Db       6021 CGGCGTACGTGCCCTTCACGCCGAAGGGGCACGTGTCTTTCTCGAGCTCGGGCCTCACGC 6080

Qy        828 CACCCTCACCACCCTCACCCACCACAACCTCCCCAACCCCCCCACCACCACCCTCACCCT 887
              |||| || ||||  | ||  ||  ||||     |   |     |||  | |    |
Db       6081 TGTCCTCTCCGCCCTTGCGCAAGACGCCCTCGGACAGGACGAAGGCACGTCGCCATGCGC 6140

Qy        888 CACCCACCCCCCACCACCAC 906
              |  ||  |||  || || |
Db       6141 CTTCCTTCCCACCCTCCGC 6159


RESULT 3
US-08-457-342-6
; Sequence 6, Application US/08457342
; Patent No. 5662898
;  GENERAL INFORMATION:
;    APPLICANT:  Schupp, Thomas
;    APPLICANT:  Ligon, James M.
;    APPLICANT:  Beck, James Joseph
;    APPLICANT:  Hill, Dwight Steven
;    APPLICANT:  Ryals, John Andrew
;    APPLICANT:  Gaffney, Thomas Deane
;    APPLICANT:  Lam, Stephen Ting
;    APPLICANT:  Hammer, Phillip E.
;    APPLICANT:  Uknes, Scott Joseph
;    TITLE OF INVENTION:  Genes for the synthesis of
;    TITLE OF INVENTION:  antipathogenic substances
;    NUMBER OF SEQUENCES:  22
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Ciba-Geigy Corporation
;      STREET:  7 Skyline Drive
;      CITY:  Hawthorne
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10532
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/457,342
;      FILING DATE:  01-JUN-1995
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/457,205
;      FILING DATE:  01-JUN-1995
;      APPLICATION NUMBER:  08/258,261
;      FILING DATE:  08-Jun-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Elmer, James Scott
;      REGISTRATION NUMBER:  36,129
;      REFERENCE/DOCKET NUMBER:  CGC 1506/CIP3
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-541-8614
;      TELEFAX:  919-541-8689
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  28958 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
US-08-457-342-6

  Query Match             28.8%;  Score 266.2;  DB 1;  Length 28958;
  Best Local Similarity   59.0%;  Pred. No. 6.4e-37;
  Matches  542;  Conservative    0;  Mismatches  353;  Indels   24;  Gaps    4;

Qy          3 CGCATTCATCTGCTCCGGACAGGGCACCCAACGCCCCGGCATGGCCCACGGCCTCTACCA 62
              |||  || ||| | | || || || | ||| || |||  ||||| || ||||||||||  |
Db       5250 CGCCGTCCTCTTCACGGGGCAAGGAAGCCAGCGGCCCACCATGGGCCGCGGCCTCTACGA 5309

Qy         63 CACCCACCCCGTCTTCGCCGCCGCACTCAACGACATCTGCACCCACCTCGACCCCCACCT 122
              |     ||||||||||   | ||| ||| ||  | || || |||||||||||| | | ||
Db       5310 CGTTTTCCCCGTCTTCCGGGACGCCCTCGACACCGTCGGCGCCCACCTCGACCGCGAGCT 5369

Qy        123 CGACCACCCCCTCCTCCCCCTCCTCACCCAAAACGACAACGACAACGAGGACGCGGCCGC 182
              ||||| |||||| | |  | |||||          ||    ||||  |   ||  ||||||
Db       5370 CGACCGCCCCCTGCGCGACGTCCTC------TTCGCTCCCGACGGCTCCGAGCAGGCCGC 5423

Qy        183 ACTGCTCCAGCAGACCCGCTACGCCCAGCCCGCCCTCTTCGCCTTCCAGGTCGCCCTCCA 242
              |  ||| |||| |||  || | |||||| ||||| || ||| || | |||||||||
Db       5424 GCGCCTCGAGCAAACCGCCTTCACCCAGCCGGCCCTGTTTGCCCTCGAAGTCGCCCTCTT 5483

Qy        243 CCGCCTCCTCACCGACGGCTACCACATCACCCCCCACTACTACGCCGGACACTCCCTCGG 302
              |  || ||     |   || |    | |  |||      |  |  |||||||||| | ||
Db       5484 TCAGCTTCT---ACAATCCTTCGGTCTGAAGCCCGCTCTCCTCCTCGGACACTCCATTGG 5540

Qy        303 CGAAATCACCGCCGCCCACCTCGCCGGCATCCTCACCCTCACCGACGCCACCACCCTCAT 362
              |||  ||  |||||||||| |||||||| ||||  | |||    |||| |  |||||||
Db       5541 CGAGCTCGTCGCCGCCCACGTCGCCGGCGTCCTTTCTCTCCAGGACGGCTGCACCCTCGT 5600

Qy        363 CACCCAACGCGCCACCCTCATGCAAACCATGCCCC---CCGGCACCATGACCACCCTCCA 419
              | ||   |||||  |  |||||||||  |  | || |    |||| |||||  ||||||||
Db       5601 CGCCGCCCGCGCAAAGCTCATGCAAGCGCTCCCACAAGGCGGCGCCATGGTCACCCTCCG 5660

Qy        420 CACCACCCCCCACCACATCACCCACCACCTCACCGCCCACGAAAACGACCTCGCCATCGC 479
              ||  ||    |  |  ||  | |||  |||     || |||||  |        | ||||
Db       5661 AGCCTCCGAGGAGGAAGTCCGCGACCTTCTCCAGCCCTACGAAGGCCGAGCTAGCCTCGC 5720

Qy        480 CGCCATCAACACCCCCACCTCCCTCGTCATCAGCGGCACCCCCCACACCGTCCAACACAT 539
              ||||  ||||     ||   ||||  |||| ||   |||       || | ||      | ||
```

```
Db       5721 CGCCCTCAATGGGCCTCTCTCCACCGTCGTCGCTGGCGATGAAGACGCGGTGGTGGAGAT 5780

Qy        540 CACCACCCTCTGCCAACAACAAGGCATCAAAACCAAAACCCTCCCCACCAACCACGCCTT 599
              | ||  ||    | ||   |  ||    || |||| |  ||| | |  || ||||||||||
Db       5781 CGCCCGCCAGGCCGAAGCCCTCGGACGAAAGACCACACGCCTGCGCGTCAGCCACGCCTT 5840

Qy        600 CCACTCCCCCCACACCAACCCCATCCTCAACCAACTCCACCAGCACACCCAAACCCTCAC 659
              ||| ||||| ||||   ||   || ||| || |  ||| ||    | |||| | ||||||
Db       5841 CCATTCCCCGCACATGGACGGAATGCTCGACGACTTCCGCCGCGTCGCCCAGAGCCTCAC 5900

Qy        660 CTACCACCCACCCCACACCCCCCTCATCAC-----------CGCCAACACCCCACCCGA 707
              |||||| ||  | | || |||| ||||| |            |||   | || |   | |
Db       5901 CTACCATCCCGCACGCATCCCCATCATCTCCAACGTCACCGGCGCGCGCGCCACGGACCA 5960

Qy        708 CCAACTCCTCACCCCCCACTACTGGACCCAACAAGCCCGCAACACCGTCGACTACGCCAC 767
              | | |||  | | ||| ||||||||  ||  || |  ||| ||||||||  || | |
Db       5961 CGAGCTCGCCTCGCCCGACTACTGGGTCCGCCACGTTCGCCACACCGTCCGCTTCCTCGA 6020

Qy        768 CACCACCCAAACCCTCCACCAACACGGCGTCACCACCTACATCGAACTCGGACCCGACAA 827
              | |   |   |||| ||| |  || |       ||   |||| ||||| ||  || | |
Db       6021 CGGCGTACGTGCCCTTCACGCCGAAGGGGCACGTGTCTTTCTCGAGCTCGGGCCTCACGC 6080

Qy        828 CACCCTCACCACCCTCACCCACCACAACCTCCCCAACCCCCCCACCACCACCCTCACCCT 887
              ||||  ||  ||||  | ||  ||  ||||    |   |    |||  | |    |
Db       6081 TGTCCTCTCCGCCCTTGCGCAAGACGCCCTCGGACAGGACGAAGGCACGTCGCCATGCGC 6140

Qy        888 CACCCACCCCCCACCACCAC 906
              |  ||  |||  || || |
Db       6141 CTTCCTTCCCACCCTCCGC 6159


RESULT 4
US-08-457-646A-6
; Sequence 6, Application US/08457646A
; Patent No. 5679560
;  GENERAL INFORMATION:
;    APPLICANT:  Schupp, Thomas
;    APPLICANT:  Ligon, James M.
;    APPLICANT:  Beck, James Joseph
;    APPLICANT:  Hill, Dwight Steven
;    APPLICANT:  Ryals, John Andrew
;    APPLICANT:  Gaffney, Thomas Deane
;    APPLICANT:  Lam, Stephen Ting
;    APPLICANT:  Hammer, Phillip E.
;    APPLICANT:  Uknes, Scott Joseph
;    TITLE OF INVENTION:  Genes for the synthesis of
;    TITLE OF INVENTION:  antipathogenic substances
;    NUMBER OF SEQUENCES:  22
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Ciba-Geigy Corporation
;      STREET:  7 Skyline Drive
;      CITY:  Hawthorne
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10532
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/457,646A
;      FILING DATE:  01-JUN-1995
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/457,205
;      FILING DATE:  01-JUN-1995
;      APPLICATION NUMBER:  08/258,261
;      FILING DATE:  08-Jun-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Elmer, James Scott
;      REGISTRATION NUMBER:  36,129
;      REFERENCE/DOCKET NUMBER:  CGC 1506/CIP3
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-541-8614
;      TELEFAX:  919-541-8689
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  28958 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
US-08-457-646A-6

  Query Match             28.8%;  Score 266.2;  DB 1;  Length 28958;
  Best Local Similarity   59.0%;  Pred. No. 6.4e-37;
  Matches  542;  Conservative    0;  Mismatches  353;  Indels   24;  Gaps    4;

Qy          3 CGCATTCATCTGCTCCGGACAGGGCACCCAACGCCCCGGCATGGCCCACGGCCTCTACCA 62
              ||| || ||| | | || || || | ||| || |||  ||||| || |||||||||| |
Db       5250 CGCCGTCCTCTTCACGGGGCAAGGAAGCCAGCGGCCCACCATGGGCCGCGGCCTCTACGA 5309

Qy         63 CACCCACCCCGTCTTCGCCGCCGCACTCAACGACATCTGCACCCACCTCGACCCCCACCT 122
              |     |||||||||    | ||| ||| ||  | || || |||||||||||| | | ||
Db       5310 CGTTTTCCCCGTCTTCCGGGACGCCCTCGACACCGTCGGCGCCCACCTCGACCGCGAGCT 5369

Qy        123 CGACCACCCCCTCCTCCCCCTCCTCACCCAAAACGACAACGACAACGAGGACGCGGCCGC 182
              ||||| |||||| | |  | |||||        ||    ||||  |   ||   ||||||
Db       5370 CGACCGCCCCCTGCGCGACGTCCTC------TTCGCTCCCGACGGCTCCGAGCAGGCCGC 5423

Qy        183 ACTGCTCCAGCAGACCCGCTACGCCCAGCCCGCCCTCTTCGCCTTCCAGGTCGCCCTCCA 242
              |  ||| |||| |||  || | ||||||| ||||| || ||| || | |||||||||
Db       5424 GCGCCTCGAGCAAACCGCCTTCACCCAGCCGGCCCTGTTTGCCCTCGAAGTCGCCCTCTT 5483

Qy        243 CCGCCTCCTCACCGACGGCTACCACATCACCCCCCACTACTACGCCGGACACTCCCTCGG 302
              |  || ||    |  || |  | |  |||    | |  |||||||||| | ||
Db       5484 TCAGCTTCT---ACAATCCTTCGGTCTGAAGCCCGCTCTCCTCCTCGGACACTCCATTGG 5540

Qy        303 CGAAATCACCGCCGCCCACCTCGCCGGCATCCTCACCCTCACCGACGCCACCACCCTCAT 362
              |||  || |||||||||| |||||||| ||||  | |||   |||| |  |||||| |
Db       5541 CGAGCTCGTCGCCGCCCACGTCGCCGGCGTCCTTTCTCTCCAGGACGGCTGCACCCTCGT 5600

Qy        363 CACCCAACGCGCCACCCTCATGCAAACCATGCCCC---CCGGCACCATGACCACCCTCCA 419
              | ||  ||||| |  |||||||||  |  || |    |||| |||||  |||||||
Db       5601 CGCCGCCCGCGCAAAGCTCATGCAAGCGCTCCCACAAGGCGGCGCCATGGTCACCCTCCG 5660

Qy        420 CACCACCCCCCACCACATCACCCACCACCTCACCGCCCACGAAAACGACCTCGCCATCGC 479
              | ||   |  | ||  | |||  |||     || ||||  |        | ||||
Db       5661 AGCCTCCGAGGAGGAAGTCCGCGACCTTCTCCAGCCCTACGAAGGCCGAGCTAGCCTCGC 5720

Qy        480 CGCCATCAACACCCCCACCTCCCTCGTCATCAGCGGCACCCCCCACACCGTCCAACACAT 539
              |||| ||||    ||   ||||  |||| ||   |||       || | ||     | ||
Db       5721 CGCCCTCAATGGGCCTCTCTCCACCGTCGTCGCTGGCGATGAAGACGCGGTGGTGGAGAT 5780

Qy        540 CACCACCCTCTGCCAACAACAAGGCATCAAAACCAAAACCCTCCCCACCAACCACGCCTT 599
              | ||  ||    | ||   |  ||    || |||| |  ||| | |  || ||||||||||
Db       5781 CGCCCGCCAGGCCGAAGCCCTCGGACGAAAGACCACACGCCTGCGCGTCAGCCACGCCTT 5840

Qy        600 CCACTCCCCCCACACCAACCCCATCCTCAACCAACTCCACCAGCACACCCAAACCCTCAC 659
              ||| ||||| ||||   ||   || ||| || |  ||| ||    | |||| | ||||||
Db       5841 CCATTCCCCGCACATGGACGGAATGCTCGACGACTTCCGCCGCGTCGCCCAGAGCCTCAC 5900

Qy        660 CTACCACCCACCCCACACCCCCCTCATCAC-----------CGCCAACACCCCACCCGA 707
              |||||| ||  | | || |||| ||||| |            |||   | || |   | |
Db       5901 CTACCATCCCGCACGCATCCCCATCATCTCCAACGTCACCGGCGCGCGCGCCACGGACCA 5960

Qy        708 CCAACTCCTCACCCCCCACTACTGGACCCAACAAGCCCGCAACACCGTCGACTACGCCAC 767
              | | |||  | | ||| ||||||||  ||  || |  ||| ||||||||  || | |
Db       5961 CGAGCTCGCCTCGCCCGACTACTGGGTCCGCCACGTTCGCCACACCGTCCGCTTCCTCGA 6020

Qy        768 CACCACCCAAACCCTCCACCAACACGGCGTCACCACCTACATCGAACTCGGACCCGACAA 827
              | |   |   |||| ||| |  || |       ||   |||| ||||| ||  || | |
Db       6021 CGGCGTACGTGCCCTTCACGCCGAAGGGGCACGTGTCTTTCTCGAGCTCGGGCCTCACGC 6080
```

```
Qy        828 CACCCTCACCACCCTCACCCACCACAACCTCCCCAACCCCCCCCACCACCACCCTCACCCT 887
              |||| || ||||  | ||  ||  ||||     |    |||  | |    |
Db       6081 TGTCCTCTCCGCCCTTGCGCAAGACGCCCTCGGACAGGACGAAGGCACGTCGCCATGCGC 6140

Qy        888 CACCCACCCCCCACCACCAC 906
              |  ||  |||  || || |
Db       6141 CTTCCTTCCCACCCTCCGC 6159


RESULT 5
US-08-458-076A-6
; Sequence 6, Application US/08458076A
; Patent No. 5698425
;  GENERAL INFORMATION:
;    APPLICANT:  Schupp, Thomas
;    APPLICANT:  Ligon, James M.
;    APPLICANT:  Beck, James Joseph
;    APPLICANT:  Hill, Dwight Steven
;    APPLICANT:  Ryals, John Andrew
;    APPLICANT:  Gaffney, Thomas Deane
;    APPLICANT:  Lam, Stephen Ting
;    APPLICANT:  Hammer, Phillip E.
;    APPLICANT:  Uknes, Scott Joseph
;    TITLE OF INVENTION:  Genes for the synthesis of
;    TITLE OF INVENTION:  antipathogenic substances
;    NUMBER OF SEQUENCES:  22
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Ciba-Geigy Corporation
;      STREET:  7 Skyline Drive
;      CITY:  Hawthorne
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10532
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/458,076A
;      FILING DATE:  01-JUN-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/457,205
;      FILING DATE:  01-JUN-1995
;      APPLICATION NUMBER:  08/258,261
;      FILING DATE:  08-Jun-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Elmer, James Scott
;      REGISTRATION NUMBER:  36,129
;      REFERENCE/DOCKET NUMBER:  CGC 1506/CIP3
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-541-8614
;      TELEFAX:  919-541-8689
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  28958 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
US-08-458-076A-6

  Query Match             28.8%;  Score 266.2;  DB 1;  Length 28958;
  Best Local Similarity   59.0%;  Pred. No. 6.4e-37;
  Matches  542;  Conservative    0;  Mismatches  353;  Indels   24;  Gaps    4;

Qy          3 CGCATTCATCTGCTCCGGACAGGGCACCCAACGCCCCGGCATGGCCCACGGCCTCTACCA 62
              |||  || ||| | | || || || | ||| || |||  ||||| || ||||||||||| |
Db       5250 CGCCGTCCTCTTCACGGGGCAAGGAAGCCAGCGGCCCACCATGGGCCGCGGCCTCTACGA 5309

Qy         63 CACCCACCCCGTCTTCGCCGCCGCACTCAACGACATCTGCACCCACCTCGACCCCCACCT 122
              |     |||||||||||    | ||| ||| ||  | || || ||||||||||||| | | ||
Db       5310 CGTTTTCCCCGTCTTCCGGGACGCCCTCGACACCGTCGGCGCCCACCTCGACCGCGAGCT 5369

Qy        123 CGACCACCCCCTCCTCCCCCTCCTCACCCAAAACGACAACGACAACGAGGACGCGGCCGC 182
              ||||| |||||| | | | |||||          ||    |||| |   ||   ||||||
Db       5370 CGACCGCCCCCTGCGCGACGTCCTC------TTCGCTCCCGACGGCTCCGAGCAGGCCGC 5423

Qy        183 ACTGCTCCAGCAGACCCGCTACGCCCAGCCCGCCCTCTTCGCCTTCCAGGTCGCCCTCCA 242
              |  ||| |||| |||  || | ||||||| ||||| || ||| || | |||||||||
Db       5424 GCGCCTCGAGCAAACCGCCTTCACCCAGCCGGCCCTGTTTGCCCTCGAAGTCGCCCTCTT 5483

Qy        243 CCGCCTCCTCACCGACGGCTACCACATCACCCCCCACTACTACGCCGGACACTCCCTCGG 302
              | || ||     |   || |   | |  |||      |  | ||||||||||| | ||
Db       5484 TCAGCTTCT---ACAATCCTTCGGTCTGAAGCCCGCTCTCCTCCTCGGACACTCCATTGG 5540

Qy        303 CGAAATCACCGCCGCCCACCTCGCCGGCATCCTCACCCTCACCGACGCCACCACCCTCAT 362
              ||| || |||||||||| |||||||| |||| | |||   |||| |  |||||| |
Db       5541 CGAGCTCGTCGCCGCCCACGTCGCCGGCGTCCTTTCTCTCCAGGACGGCTGCACCCTCGT 5600

Qy        363 CACCCAACGCGCCACCCTCATGCAAACCATGCCCC---CCGGCACCATGACCACCCTCCA 419
              | ||   ||||| |  |||||||| |  | || |    |||| |||||  ||||||||
Db       5601 CGCCGCCCGCGCAAAGCTCATGCAAGCGCTCCCACAAGGCGGCGCCATGGTCACCCTCCG 5660

Qy        420 CACCACCCCCCACCACATCACCCACCACCTCACCGCCCACGAAAACGACCTCGCCATCGC 479
              || ||    |  |  ||  | |||  |||     || |||||  |          | ||||
Db       5661 AGCCTCCGAGGAGGAAGTCCGCGACCTTCTCCAGCCCTACGAAGGCCGAGCTAGCCTCGC 5720

Qy        480 CGCCATCAACACCCCCACCTCCCTCGTCATCAGCGGCACCCCCCACACCGTCCAACACAT 539
              |||| ||||    ||   ||||  |||| ||    |||         || | ||     | ||
Db       5721 CGCCCTCAATGGGCCTCTCTCCACCGTCGTCGCTGGCGATGAAGACGCGGTGGTGGAGAT 5780

Qy        540 CACCACCCTCTGCCAACAACAAGGCATCAAAACCAAAACCCTCCCCACCAACCACGCCTT 599
              | ||  ||    | ||   |  ||    || |||| |  ||| | |  || |||||||||
Db       5781 CGCCCGCCAGGCCGAAGCCCTCGGACGAAAGACCACACGCCTGCGCGTCAGCCACGCCTT 5840

Qy        600 CCACTCCCCCCACACCAACCCCATCCTCAACCAACTCCACCAGCACACCCAAACCCTCAC 659
              ||| ||||| ||||   ||   || ||| || |  ||| ||     | |||| | ||||||
Db       5841 CCATTCCCCGCACATGGACGGAATGCTCGACGACTTCCGCCGCGTCGCCCAGAGCCTCAC 5900

Qy        660 CTACCACCCACCCCACACCCCCCTCATCAC-----------CGCCAACACCCCACCCGA 707
              |||||| ||  | | || |||| ||||| |           |||  | || |  | |
Db       5901 CTACCATCCCGCACGCATCCCCATCATCTCCAACGTCACCGGCGCGCGCGCCACGGACCA 5960

Qy        708 CCAACTCCTCACCCCCCACTACTGGACCCAACAAGCCCGCAACACCGTCGACTACGCCAC 767
              | | |||  | | ||| |||||||  ||  || |  |||  |||||||  || |  |
Db       5961 CGAGCTCGCCTCGCCCGACTACTGGGTCCGCCACGTTCGCCACACCGTCCGCTTCCTCGA 6020

Qy        768 CACCACCCAAACCCTCCACCAACACGGCGTCACCACCTACATCGAACTCGGACCCGACAA 827
              | |   |  |||| |||    | || |       ||   |||| ||||| ||  ||
Db       6021 CGGCGTACGTGCCCTTCACGCCGAAGGGGCACGTGTCTTTCTCGAGCTCGGGCCTCACGC 6080

Qy        828 CACCCTCACCACCCTCACCCACCACAACCTCCCCAACCCCCCCACCACCACCCTCACCCT 887
              |||| || ||||  | ||  ||  ||||     |    |||  | |    |
Db       6081 TGTCCTCTCCGCCCTTGCGCAAGACGCCCTCGGACAGGACGAAGGCACGTCGCCATGCGC 6140

Qy        888 CACCCACCCCCCACCACCAC 906
              |  ||  |||  || || |
Db       6141 CTTCCTTCCCACCCTCCGC 6159


RESULT 6
US-08-764-233A-4
; Sequence 4, Application US/08764233A
; Patent No. 5716849
;  GENERAL INFORMATION:
;    APPLICANT:  Ligon, James M.
;    APPLICANT:  Schupp, Thomas
;    APPLICANT:  Beck, James J.
;    APPLICANT:  Hill, Dwight S.
;    APPLICANT:  Neff, Snezanna
;    APPLICANT:  Ryals, John A.
```

```
;    TITLE OF INVENTION:  Genes For The Biosynthesis Of Soraphen
;    NUMBER OF SEQUENCES:  10
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Ciba-Geigy Corporation
;      STREET:  520 White Plains Road, P.O. Box 2005
;      CITY:  Tarrytown
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10591
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/764,233A
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/729,214
;      FILING DATE:  09-OCT-1996
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/258,261
;      FILING DATE:  08-JUN-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Meigs, J. Timothy
;      REGISTRATION NUMBER:  38,241
;      REFERENCE/DOCKET NUMBER:  1506/CIP6
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (919) 541-8587
;      TELEFAX:  (919) 541-8689
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  28958 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    ORIGINAL SOURCE:
;      ORGANISM:  Sorangium cellulosum
;    IMMEDIATE SOURCE:
;      CLONE:  p98/1
US-08-764-233A-4

  Query Match             28.8%;  Score 266.2;  DB 1;  Length 28958;
  Best Local Similarity   59.0%;  Pred. No. 6.4e-37;
  Matches  542;  Conservative    0;  Mismatches  353;  Indels   24;  Gaps    4;

Qy          3 CGCATTCATCTGCTCCGGACAGGGCACCCAACGCCCCGGCATGGCCCACGGCCTCTACCA 62
              ||| || ||| | | || || || | ||| || |||  ||||| || ||||||||||| |
Db       5250 CGCCGTCCTCTTCACGGGGCAAGGAAGCCAGCGGCCCACCATGGGCCGCGGCCTCTACGA 5309

Qy         63 CACCCACCCCGTCTTCGCCGCCGCACTCAACGACATCTGCACCCACCTCGACCCCCACCT 122
              |     ||||||||||   | ||| ||| ||  | || || ||||||||||| | | ||
Db       5310 CGTTTTCCCCGTCTTCCGGGACGCCCTCGACACCGTCGGCGCCCACCTCGACCGCGAGCT 5369

Qy        123 CGACCACCCCCTCCTCCCCCTCCTCACCCAAAACGACAACGACAACGAGGACGCGGCCGC 182
              ||||| |||||| | |  | |||||        ||    ||||  |   ||  ||||||
Db       5370 CGACCGCCCCCTGCGCGACGTCCTC------TTCGCTCCCGACGGCTCCGAGCAGGCCGC 5423

Qy        183 ACTGCTCCAGCAGACCCGCTACGCCCAGCCCGCCCTCTTCGCCTTCCAGGTCGCCCTCCA 242
              |  ||| |||| |||  || | ||||||| ||||| || ||| || | |||||||||
Db       5424 GCGCCTCGAGCAAACCGCCTTCACCCAGCCGGCCCTGTTTGCCCTCGAAGTCGCCCTCTT 5483

Qy        243 CCGCCTCCTCACCGACGGCTACCACATCACCCCCCACTACTACGCCGGACACTCCCTCGG 302
              |  || ||    |  || |    | |  |||    |  |  |||||||||| | ||
Db       5484 TCAGCTTCT---ACAATCCTTCGGTCTGAAGCCCGCTCTCCTCCTCGGACACTCCATTGG 5540

Qy        303 CGAAATCACCGCCGCCCACCTCGCCGGCATCCTCACCCTCACCGACGCCACCACCCTCAT 362
              |||  || ||||||||||| |||||||| |||| | |||   |||| |  ||||||| |
Db       5541 CGAGCTCGTCGCCGCCCACGTCGCCGGCGTCCTTTCTCTCCAGGACGGCTGCACCCTCGT 5600

Qy        363 CACCCAACGCGCCACCCTCATGCAAACCATGCCCC---CCGGCACCATGACCACCCTCCA 419
              |  ||   ||||| |  ||||||||| |  | || |    |||| ||||| ||||||||
Db       5601 CGCCGCCCGCGCAAAGCTCATGCAAGCGCTCCCACAAGGCGGCGCCATGGTCACCCTCCG 5660

Qy        420 CACCACCCCCCACCACATCACCCACCACCTCACCGCCCACGAAAACGACCTCGCCATCGC 479
              || ||    |  |  || | |||  |||    || ||||  |         | ||||
Db       5661 AGCCTCCGAGGAGGAAGTCCGCGACCTTCTCCAGCCCTACGAAGGCCGAGCTAGCCTCGC 5720

Qy        480 CGCCATCAACACCCCCACCTCCCTCGTCATCAGCGGCACCCCCCACACCGTCCAACACAT 539
              |||| ||||    ||   ||||  |||| ||   |||        || | ||     | ||
Db       5721 CGCCCTCAATGGGCCTCTCTCCACCGTCGTCGCTGGCGATGAAGACGCGGTGGTGGAGAT 5780

Qy        540 CACCACCCTCTGCCAACAACAAGGCATCAAAACCAAAACCCTCCCCACCAACCACGCCTT 599
              | ||  ||    | ||   |  ||    || |||| |  ||| | |  || ||||||||||
Db       5781 CGCCCGCCAGGCCGAAGCCCTCGGACGAAAGACCACACGCCTGCGCGTCAGCCACGCCTT 5840

Qy        600 CCACTCCCCCCACACCAACCCCATCCTCAACCAACTCCACCAGCACACCCAAACCCTCAC 659
              ||| ||||| ||||    ||   || ||| || |- ||| ||   | |||| | ||||||
Db       5841 CCATTCCCCGCACATGGACGGAATGCTCGACGACTTCCGCCGCGTCGCCCAGAGCCTCAC 5900

Qy        660 CTACCACCCACCCCACACCCCCCTCATCAC----------CGCCAACACCCCACCCGA 707
              |||||| ||  | | || |||| ||||| |           |||  | || |   | |
Db       5901 CTACCATCCCGCACGCATCCCCATCATCTCCAACGTCACCGGCGCGCGCGCCACGGACCA 5960

Qy        708 CCAACTCCTCACCCCCCACTACTGGACCCAACAAGCCCGCAACACCGTCGACTACGCCAC 767
              | | |||  | | ||| ||||||||  ||  || |  ||| ||||||||  || |  |
Db       5961 CGAGCTCGCCTCGCCCGACTACTGGGTCCGCCACGTTCGCCACACCGTCCGCTTCCTCGA 6020

Qy        768 CACCACCCAAACCCTCCACCAACACGGCGTCACCACCTACATCGAACTCGGACCCGACAA 827
              |  |   |  |||| |||    | || |         ||   |||| ||||| ||  ||
Db       6021 CGGCGTACGTGCCCTTCACGCCGAAGGGGCACGTGTCTTTCTCGAGCTCGGGCCTCACGC 6080

Qy        828 CACCCTCACCACCCTCACCCACCACAACCTCCCCAACCCCCCCACCACCACCCTCACCCT 887
              |||| || ||||  | ||  ||  ||||   |   |     ||| | |   |
Db       6081 TGTCCTCTCCGCCCTTGCGCAAGACGCCCTCGGACAGGACGAAGGCACGTCGCCATGCGC 6140

Qy        888 CACCCACCCCCCACCACCAC 906
              |  ||  |||  || || |
Db       6141 CTTCCTTCCCACCCTCCGC 6159


RESULT 7
US-08-457-335A-6
; Sequence 6, Application US/08457335A
; Patent No. 5723759
;  GENERAL INFORMATION:
;    APPLICANT:  Schupp, Thomas
;    APPLICANT:  Ligon, James M.
;    APPLICANT:  Beck, James Joseph
;    APPLICANT:  Hill, Dwight Steven
;    APPLICANT:  Ryals, John Andrew
;    APPLICANT:  Gaffney, Thomas Deane
;    APPLICANT:  Lam, Stephen Ting
;    APPLICANT:  Hammer, Phillip E.
;    APPLICANT:  Uknes, Scott Joseph
;    TITLE OF INVENTION:  Genes for the synthesis of
;    TITLE OF INVENTION:  antipathogenic substances
;    NUMBER OF SEQUENCES:  22
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Ciba-Geigy Corporation
;      STREET:  7 Skyline Drive
;      CITY:  Hawthorne
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10532
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/457,335A
;      FILING DATE:  01-JUN-1995
;      CLASSIFICATION:  800
```

```
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/457,205
;      FILING DATE:  01-JUN-1995
;      APPLICATION NUMBER:  08/258,261
;      FILING DATE:  08-Jun-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Elmer, James Scott
;      REGISTRATION NUMBER:  36,129
;      REFERENCE/DOCKET NUMBER:  CGC 1506/CIP3
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-541-8614
;      TELEFAX:  919-541-8689
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  28958 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
US-08-457-335A-6

  Query Match           28.8%;  Score 266.2;  DB 1;  Length 28958;
  Best Local Similarity   59.0%;  Pred. No. 6.4e-37;
  Matches  542;  Conservative    0;  Mismatches  353;  Indels   24;  Gaps    4;

Qy          3 CGCATTCATCTGCTCCGGACAGGGCACCCAACGCCCCGGCATGGCCCACGGCCTCTACCA 62
              |||  || ||| | | || || || | ||| || |||  ||||| || ||||||||||
Db       5250 CGCCGTCCTCTTCACGGGGCAAGGAAGCCAGCGGCCCACCATGGGCCGCGGCCTCTACGA 5309

Qy         63 CACCCACCCCGTCTTCGCCGCCGCACTCAACGACATCTGCACCCACCTCGACCCCCACCT 122
              |     ||||||||||    | ||| ||| ||  | || || ||||||||||||| | | ||
Db       5310 CGTTTTCCCCGTCTTCCGGGACGCCCTCGACACCGTCGGCGCCCACCTCGACCGCGAGCT 5369

Qy        123 CGACCACCCCCTCCTCCCCCTCCTCACCCAAAACGACAACGACAACGAGGACGCGGCCGC 182
              ||||| |||||| | |  | |||||         ||    |||| |   ||   ||||||
Db       5370 CGACCGCCCCCTGCGCGACGTCCTC------TTCGCTCCCGACGGCTCCGAGCAGGCCGC 5423

Qy        183 ACTGCTCCAGCAGACCCGCTACGCCCAGCCCGCCCTCTTCGCCTTCCAGGTCGCCCTCCA 242
              |  ||| |||| |||  || | ||||||| ||||| || ||| || | ||||||||
Db       5424 GCGCCTCGAGCAAACCGCCTTCACCCAGCCGGCCCTGTTTGCCCTCGAAGTCGCCCTCTT 5483

Qy        243 CCGCCTCCTCACCGACGGCTACCACATCACCCCCCACTACTACGCCGGACACTCCCTCGG 302
              |  || ||     |   || |    | |  |||      | |  |||||||||| | ||
Db       5484 TCAGCTTCT---ACAATCCTTCGGTCTGAAGCCCGCTCTCCTCCTCGGACACTCCATTGG 5540

Qy        303 CGAAATCACCGCCGCCCACCTCGCCGGCATCCTCACCCTCACCGACGCCACCACCCTCAT 362
              |||  ||  |||||||||| |||||||| ||||  | |||   |||| |  ||||||| |
Db       5541 CGAGCTCGTCGCCGCCCACGTCGCCGGCGTCCTTTCTCTCCAGGACGGCTGCACCCTCGT 5600

Qy        363 CACCCAACGCGCCACCCTCATGCAAACCATGCCCC---CCGGCACCATGACCACCCTCCA 419
              | ||   ||||| |  |||||||||  |  | || |    |||| |||||  ||||||||
Db       5601 CGCCGCCCGCGCAAAGCTCATGCAAGCGCTCCCACAAGGCGGCGCCATGGTCACCCTCCG 5660

Qy        420 CACCACCCCCCACCACATCACCCACCACCTCACCGCCCACGAAAACGACCTCGCCATCGC 479
              || ||    |  |  ||  | |||  |||    || |||||  |        | ||||
Db       5661 AGCCTCCGAGGAGGAAGTCCGCGACCTTCTCCAGCCCTACGAAGGCCGAGCTAGCCTCGC 5720

Qy        480 CGCCATCAACACCCCCACCTCCCTCGTCATCAGCGGCACCCCCCACACCGTCCAACACAT 539
              ||||  ||||    ||   ||||  |||| ||   |||        || | ||     | ||
Db       5721 CGCCCTCAATGGGCCTCTCTCCACCGTCGTCGCTGGCGATGAAGACGCGGTGGTGGAGAT 5780

Qy        540 CACCACCCTCTGCCAACAACAAGGCATCAAAACCAAAACCCTCCCCACCAACCACGCCTT 599
              | ||  ||    | ||   |  ||    || |||| |  ||| | |  || ||||||||||
Db       5781 CGCCCGCCAGGCCGAAGCCCTCGGACGAAAGACCACACGCCTGCGCGTCAGCCACGCCTT 5840

Qy        600 CCACTCCCCCCACACCAACCCCATCCTCAACCAACTCCACCAGCACACCCAAACCCTCAC 659
              ||| ||||| ||||   ||   || ||| || |  ||| ||     | |||| | ||||||
Db       5841 CCATTCCCCGCACATGGACGGAATGCTCGACGACTTCCGCCGCGTCGCCCAGAGCCTCAC 5900

Qy        660 CTACCACCCACCCCACACCCCCCTCATCAC-----------CGCCAACACCCCACCCGA 707
              |||||| ||  | | || |||| ||||| |           |||   | || |   | |
```

```
Db       5901 CTACCATCCCGCACGCATCCCCATCATCTCCAACGTCACCGGCGCGCGCGCCACGGACCA 5960

Qy        708 CCAACTCCTCACCCCCCACTACTGGACCCAACAAGCCCGCAACACCGTCGACTACGCCAC 767
              | | |||  | | ||| ||||||||  ||  || |  ||| ||||||||  || | |
Db       5961 CGAGCTCGCCTCGCCCGACTACTGGGTCCGCCACGTTCGCCACACCGTCCGCTTCCTCGA 6020

Qy        768 CACCACCCAAACCCTCCACCAACACGGCGTCACCACCTACATCGAACTCGGACCCGACAA 827
              |  |   |  |||| |||    | || |       ||   |||| ||||| ||  ||
Db       6021 CGGCGTACGTGCCCTTCACGCCGAAGGGGCACGTGTCTTTCTCGAGCTCGGGCCTCACGC 6080

Qy        828 CACCCTCACCACCCTCACCCACCACAACCTCCCCAACCCCCCCACCACCACCCTCACCCT 887
              ||||| || ||||  | ||  ||  ||||     |   |      ||| || |     |
Db       6081 TGTCCTCTCCGCCCTTGCGCAAGACGCCCTCGGACAGGACGAAGGCACGTCGCCATGCGC 6140

Qy        888 CACCCACCCCCACCACCAC 906
              | ||  |||  || || |
Db       6141 CTTCCTTCCCACCCTCCGC 6159


RESULT 8
US-08-729-214-6
; Sequence 6, Application US/08729214
; Patent No. 5817502
;  GENERAL INFORMATION:
;    APPLICANT:  Ligon, James M.
;    APPLICANT:  Hill, Dwight Steven
;    APPLICANT:  Ryals, John Andrew
;    APPLICANT:  Hammer, Phillip E.
;    APPLICANT:  van Pee, Karl-Heinz
;    APPLICANT:  Kirner, Sabine
;    TITLE OF INVENTION:  Genes for the synthesis of
;    TITLE OF INVENTION:  antipathogenic substances
;    NUMBER OF SEQUENCES:  27
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Ciba-Geigy Corporation
;      STREET:  520 White Plains Road
;      CITY:  Tarrytown
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10591
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/729,214
;      FILING DATE:  TBA
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Meigs, J. Timothy
;      REGISTRATION NUMBER:  38,241
;      REFERENCE/DOCKET NUMBER:  CGC 1506/CIP5
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-541-8587
;      TELEFAX:  919-541-8689
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  28958 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
US-08-729-214-6

  Query Match           28.8%;  Score 266.2;  DB 1;  Length 28958;
  Best Local Similarity   59.0%;  Pred. No. 6.4e-37;
  Matches  542;  Conservative    0;  Mismatches  353;  Indels   24;  Gaps    4;

Qy          3 CGCATTCATCTGCTCCGGACAGGGCACCCAACGCCCCGGCATGGCCCACGGCCTCTACCA 62
              |||  || ||| | | || || || | ||| || |||  ||||| || ||||||||||
```

```
Db       5250 CGCCGTCCTCTTCACGGGGCAAGGAAGCCAGCGGCCCACCATGGGCCGCGGCCTCTACGA 5309

Qy         63 CACCCACCCCGTCTTCGCCGCCGCACTCAACGACATCTGCACCCACCTCGACCCCCACCT 122

Db       5310 CGTTTTCCCCGTCTTCCGGGACGCCCTCGACACCGTCGGCGCCCACCTCGACCGCGAGCT 5369

Qy        123 CGACCACCCCCTCCTCCCCCTCCTCACCCAAAACGACAACGACAACGAGGACGCGGCCGC 182

Db       5370 CGACCGCCCCCTGCGCGACGTCCTC------TTCGCTCCCGACGGCTCCGAGCAGGCCGC 5423

Qy        183 ACTGCTCCAGCAGACCCGCTACGCCCAGCCCGCCCTCTTCGCCTTCCAGGTCGCCCTCCA 242

Db       5424 GCGCCTCGAGCAAACCGCCTTCACCCAGCCGGCCCTGTTTGCCCTCGAAGTCGCCCTCTT 5483

Qy        243 CCGCCTCCTCACCGACGGCTACCACATCACCCCCCACTACTACGCCGGACACTCCCTCGG 302

Db       5484 TCAGCTTCT---ACAATCCTTCGGTCTGAAGCCCGCTCTCCTCCTCGGACACTCCATTGG 5540

Qy        303 CGAAATCACCGCCGCCCACCTCGCCGGCATCCTCACCCTCACCGACGCCACCACCCTCAT 362

Db       5541 CGAGCTCGTCGCCGCCCACGTCGCCGGCGTCCTTTCTCTCCAGGACGGCTGCACCCTCGT 5600

Qy        363 CACCCAACGCGCCACCCTCATGCAAACCATGCCCC---CCGGCACCATGACCACCCTCCA 419

Db       5601 CGCCGCCCGCGCAAAGCTCATGCAAGCGCTCCCACAAGGCGGCGCCATGGTCACCCTCCG 5660

Qy        420 CACCACCCCCCACCACATCACCCACCACCTCACCGCCCACGAAAACGACCTCGCCATCGC 479

Db       5661 AGCCTCCGAGGAGGAAGTCCGCGACCTTCTCCAGCCCTACGAAGGCCGAGCTAGCCTCGC 5720

Qy        480 CGCCATCAACACCCCCACCTCCCTCGTCATCAGCGGCACCCCCCACACCGTCCAACACAT 539

Db       5721 CGCCCTCAATGGGCCTCTCTCCACCGTCGTCGCTGGCGATGAAGACGCGGTGGTGGAGAT 5780

Qy        540 CACCACCCTCTGCCAACAACAAGGCATCAAAACCAAAACCCTCCCCACCAACCACGCCTT 599

Db       5781 CGCCCGCCAGGCCGAAGCCCTCGGACGAAAGACCACACGCCTGCGCGTCAGCCACGCCTT 5840

Qy        600 CCACTCCCCCCACACCAACCCCATCCTCAACCAACTCCACCAGCACACCCAAACCCTCAC 659

Db       5841 CCATTCCCCGCACATGGACGGAATGCTCGACGACTTCCGCCGCGTCGCCCAGAGCCTCAC 5900

Qy        660 CTACCACCCACCCCACACCCCCCTCATCAC------------CGCCAACACCCCACCCGA 707

Db       5901 CTACCATCCCGCACGCATCCCCATCATCTCCAACGTCACCGGCGCGCGCGCCACGGACCA 5960

Qy        708 CCAACTCCTCACCCCCCACTACTGGACCCAACAAGCCCGCAACACCGTCGACTACGCCAC 767

Db       5961 CGAGCTCGCCTCGCCCGACTACTGGGTCCGCCACGTTCGCCACACCGTCCGCTTCCTCGA 6020

Qy        768 CACCACCCAAACCCTCCACCAACACGGCGTCACCACCTACATCGAACTCGGACCCGACAA 827

Db       6021 CGGCGTACGTGCCCTTCACGCCGAAGGGGCACGTGTCTTTCTCGAGCTCGGGCCTCACGC 6080

Qy        828 CACCCTCACCACCCTCACCCACCACAACCTCCCCAACCCCCCCACCACCACCCTCACCCT 887

Db       6081 TGTCCTCTCCGCCCTTGCGCAAGACGCCCTCGGACAGGACGAAGGCACGTCGCCATGCGC 6140

Qy        888 CACCCACCCCCCACCACCAC 906

Db       6141 CTTCCTTCCCACCCTCCGC 6159


RESULT 9
US-09-028-934-6
; Sequence 6, Application US/09028934
; Patent No. 6117670
;  GENERAL INFORMATION:
;    APPLICANT:  Ligon, James M.
;    APPLICANT:  Hill, Dwight S.
;    APPLICANT:  Lam, Steven T.
;    APPLICANT:  Hammer, Philip E.
;    APPLICANT:  van Pee, Karl-Heinz
;    APPLICANT:  Kirner, Sabine
;    APPLICANT:  Young, Thomas R.
;    TITLE OF INVENTION:  Pyrrolnitrin Biosynthesis Genes and Uses
;    TITLE OF INVENTION:  Thereof
;    NUMBER OF SEQUENCES:  37
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  No. 6117670artis Corporation
;      STREET:  3054 Cornwallis Road
;      CITY:  Research Triangle Park
;      STATE:  NC
;      COUNTRY:  USA
;      ZIP:  27709
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/028,934
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/729,214
;      FILING DATE:  09-OCT-1996
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/258,261
;      FILING DATE:  08-JUN-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Meigs, J. Timothy
;      REGISTRATION NUMBER:  38,241
;      REFERENCE/DOCKET NUMBER:  CGC1506/CIP7
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-541-8587
;      TELEFAX:  919-541-8689
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  28958 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
US-09-028-934-6

  Query Match           28.8%;  Score 266.2;  DB 3;  Length 28958;
  Best Local Similarity   59.0%;  Pred. No. 6.4e-37;
  Matches  542;  Conservative    0;  Mismatches  353;  Indels   24;  Gaps    4;

Qy          3 CGCATTCATCTGCTCCGGACAGGGCACCCAACGCCCCGGCATGGCCCACGGCCTCTACCA 62

Db       5250 CGCCGTCCTCTTCACGGGGCAAGGAAGCCAGCGGCCCACCATGGGCCGCGGCCTCTACGA 5309

Qy         63 CACCCACCCCGTCTTCGCCGCCGCACTCAACGACATCTGCACCCACCTCGACCCCCACCT 122

Db       5310 CGTTTTCCCCGTCTTCCGGGACGCCCTCGACACCGTCGGCGCCCACCTCGACCGCGAGCT 5369

Qy        123 CGACCACCCCCTCCTCCCCCTCCTCACCCAAAACGACAACGACAACGAGGACGCGGCCGC 182

Db       5370 CGACCGCCCCCTGCGCGACGTCCTC------TTCGCTCCCGACGGCTCCGAGCAGGCCGC 5423

Qy        183 ACTGCTCCAGCAGACCCGCTACGCCCAGCCCGCCCTCTTCGCCTTCCAGGTCGCCCTCCA 242

Db       5424 GCGCCTCGAGCAAACCGCCTTCACCCAGCCGGCCCTGTTTGCCCTCGAAGTCGCCCTCTT 5483

Qy        243 CCGCCTCCTCACCGACGGCTACCACATCACCCCCCACTACTACGCCGGACACTCCCTCGG 302

Db       5484 TCAGCTTCT---ACAATCCTTCGGTCTGAAGCCCGCTCTCCTCCTCGGACACTCCATTGG 5540

Qy        303 CGAAATCACCGCCGCCCACCTCGCCGGCATCCTCACCCTCACCGACGCCACCACCCTCAT 362

Db       5541 CGAGCTCGTCGCCGCCCACGTCGCCGGCGTCCTTTCTCTCCAGGACGGCTGCACCCTCGT 5600
```

```
Qy        363 CACCCAACGCGCCACCCTCATGCAAACCATGCCCC---CCGGCACCATGACCACCCTCCA 419
              | ||   ||||| |  ||||||||| |  | || |    |||| |||||  ||||||||
Db       5601 CGCCGCCCGCGCAAAGCTCATGCAAGCGCTCCCACAAGGCGGCGCCATGGTCACCCTCCG 5660

Qy        420 CACCACCCCCCACCACATCACCCACCACCTCACCGCCCACGAAAACGACCTCGCCATCGC 479
                || ||    |  |  ||  | |||  |||    || |||||  |        | ||||
Db       5661 AGCCTCCGAGGAGGAAGTCCGCGACCTTCTCCAGCCCTACGAAGGCCGAGCTAGCCTCGC 5720

Qy        480 CGCCATCAACACCCCCACCTCCCTCGTCATCAGCGGCACCCCCCACACCGTCCAACACAT 539
              |||| ||||    ||   ||||  |||| ||   |||        || | ||    | ||
Db       5721 CGCCCTCAATGGGCCTCTCTCCACCGTCGTCGCTGGCGATGAAGACGCGGTGGTGGAGAT 5780

Qy        540 CACCACCCTCTGCCAACAACAAGGCATCAAAACCAAAACCCTCCCCACCAACCACGCCTT 599
              | ||  ||    | ||   |  ||    ||.|||| |  ||| | |  || |||||||||
Db       5781 CGCCCGCCAGGCCGAAGCCCTCGGACGAAAGACCACACGCCTGCGCGTCAGCCACGCCTT 5840

Qy        600 CCACTCCCCCCACACCAACCCCATCCTCAACCAACTCCACCAGCACACCCAAACCCTCAC 659
              ||| ||||| ||||   ||   || ||| || |  ||| ||     | |||| | ||||||
Db       5841 CCATTCCCCGCACATGGACGGAATGCTCGACGACTTCCGCCGCGTCGCCCAGAGCCTCAC 5900

Qy        660 CTACCACCCACCCCACACCCCCCTCATCAC----------CGCCAACACCCCACCCGA 707
              |||||| ||  | | || |||| ||||| |          |||   | || |    | |
Db       5901 CTACCATCCCGCACGCATCCCCATCATCTCCAACGTCACCGGCGCGCGCGCCACGGACCA 5960

Qy        708 CCAACTCCTCACCCCCCACTACTGGACCCAACAAGCCCGCAACACCGTCGACTACGCCAC 767
              | | |||  |·| ||| |||||||||  ||  || |  ||| ||||||||  || | |
Db       5961 CGAGCTCGCCTCGCCCGACTACTGGGTCCGCCACGTTCGCCACACCGTCCGCTTCCTCGA 6020

Qy        768 CACCACCCAAACCCTCCACCAACACGGCGTCACCACCTACATCGAACTCGGACCCGACAA 827
              | |    |  |||| |||    | || |        ||   |||| ||||| ||  ||
Db       6021 CGGCGTACGTGCCCTTCACGCCGAAGGGGCACGTGTCTTTCTCGAGCTCGGGCCTCACGC 6080

Qy        828 CACCCTCACCACCCTCACCCACCACAACCTCCCCAACCCCCCCACCACCACCCTCACCCT 887
              |||| || ||||  | ||  ||  ||||    |   |     |||  | |   |
Db       6081 TGTCCTCTCCGCCCTTGCGCAAGACGCCCTCGGACAGGACGAAGGCACGTCGCCATGCGC 6140

Qy        888 CACCCACCCCCCACCACCAC 906
              |  ||  |||  || || |
Db       6141 CTTCCTTCCCACCCTCCGC 6159
```

```
RESULT 10
US-08-764-233A-1
; Sequence 1, Application US/08764233A
; Patent No. 5716849
;  GENERAL INFORMATION:
;    APPLICANT:  Ligon, James M.
;    APPLICANT:  Schupp, Thomas
;    APPLICANT:  Beck, James J.
;    APPLICANT:  Hill, Dwight S.
;    APPLICANT:  Neff, Snezanna
;    APPLICANT:  Ryals, John A.
;    TITLE OF INVENTION:  Genes For The Biosynthesis Of Soraphen
;    NUMBER OF SEQUENCES:  10
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Ciba-Geigy Corporation
;      STREET:  520 White Plains Road, P.O. Box 2005
;      CITY:  Tarrytown
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10591
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/764,233A
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/729,214
;      FILING DATE:  09-OCT-1996
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/258,261
;      FILING DATE:  08-JUN-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Meigs, J. Timothy
;      REGISTRATION NUMBER:  38,241
;      REFERENCE/DOCKET NUMBER:  1506/CIP6
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (919) 541-8587
;      TELEFAX:  (919) 541-8689
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  49377 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    ORIGINAL SOURCE:
;      ORGANISM:  Sorangium cellulosum
;    IMMEDIATE SOURCE:
;      CLONE:  p98/1, pJL3, and pVKM15
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  383..760
;      OTHER INFORMATION:  /product= "SorR"
;      OTHER INFORMATION:  /note= "This gene encodes a protein that is highly homologou
;      OTHER INFORMATION:  the reductase domains of type I PKSs such as eryA from
;      OTHER INFORMATION:  Saccharopolyspora erythraea."
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  927..19874
;      OTHER INFORMATION:  /product= "SorA"
;      OTHER INFORMATION:  /note= "Gene product is highly homologous to type I PKSs tha
;      OTHER INFORMATION:  are known to be involved in the synthesis of polyketide
;      OTHER INFORMATION:  compounds."
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  942..7115
;      OTHER INFORMATION:  /product= "Module 1 of SorA"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  7203..12884
;      OTHER INFORMATION:  /product= "Module 2 of SorA"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  13455..19616
;      OTHER INFORMATION:  /product= "Module 3 of SorA"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  19871..46318
;      OTHER INFORMATION:  /product= "SorB"
;      OTHER INFORMATION:  /note= "Gene product is highly homologous to type I PKS gene
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  19870..24556
;      OTHER INFORMATION:  /product= "Module 1 of SorB"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  24638..30820
;      OTHER INFORMATION:  /product= "Module 2 of SorB"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  30881..35446
;      OTHER INFORMATION:  /product= "Module 3 of SorB"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  35528..40114
;      OTHER INFORMATION:  /product= "Module 4 of SorB"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  40190..46318
;      OTHER INFORMATION:  /product= "Module 5 of SorB"
;    FEATURE:
```

```
;       NAME/KEY:  misc_feature
;       LOCATION:  46851..47891
;       OTHER INFORMATION:  /product= "SorM"
;       OTHER INFORMATION:  /note= "The protein encoded by the sorM gene is highly
;       OTHER INFORMATION:  homologous to the methyltransferase from Streptomyces
;       OTHER INFORMATION:  hygroscopicus that is involved in the synthesis of the
;       OTHER INFORMATION:  polyketide rappamicin."
US-08-764-233A-1

  Query Match            28.8%;  Score 266.2;  DB 1;  Length 49377;
  Best Local Similarity  59.0%;  Pred. No. 6.6e-37;
  Matches  542;  Conservative    0;  Mismatches  353;  Indels   24;  Gaps    4;

Qy          3 CGCATTCATCTGCTCCGGACAGGGCACCCAACGCCCCGGCATGGCCCACGGCCTCTACCA 62
              ||| || ||| | | || || || | ||| || ||| ||||| || || ||||||| |
Db       8801 CGCCGTCCTCTTCACGGGGCAAGGAAGCCAGCGGCCCACCATGGGCCGCGCCCTCTACGA 8860

Qy         63 CACCCACCCCGTCTTCGCCGCCGCACTCAACGACATCTGCACCCACCTCGACCCCCACCT 122
              | |   ||||||||||   | ||| ||| ||  | ||  | ||||||||||||| | ||||
Db       8861 CGCTTTCCCCGTCTTCCGGGACGCCCTCGACACCGTCGCCGCCCACCTCGACCGCGACCT 8920

Qy        123 CGACCACCCCCTCCTCCCCCTCCTCACCCAAAACGACAACGACAACGAGGACGCGGCCGC 182
              ||||| |||||| | |  | |||||         ||    |||| |   ||   ||||||
Db       8921 CGACCGCCCCCTGCGCGACGTCCTC------TTCGCTCCCGACGGCTCCGAGCAGGCCGC 8974

Qy        183 ACTGCTCCAGCAGACCCGCTACGCCCAGCCCGCCCTCTTCGCCTTCCAGGTCGCCCTCCA 242
              | ||| | || |||  || | ||||||| ||||| || ||| || | |||||||||
Db       8975 GCGCCTCGACCAAACCGCCTTCACCCAGCCGGCCCTGTTTGCCCTCGAAGTCGCCCTCTT 9034

Qy        243 CCGCCTCCTCACCGACGGCTACCACATCACCCCCCACTACTACGCCGGACACTCCCTCGG 302
              |  || ||   | |   || |    | |  |||     |  |  |||||||||| | ||
Db       9035 TCAGCTTCT---CCAATCCTTCGGTCTGAAGCCCGCTCTCCTCCTCGGACACTCCATTGG 9091

Qy        303 CGAAATCACCGCCGCCCACCTCGCCGGCATCCTCACCCTCACCGACGCCACCACCCTCAT 362
              |||  ||  |||||||||| |||||| ||||  | |||   ||||||  |||||||| |
Db       9092 CGAGCTCGTCGCCGCCCACGTCGCCGGCGTCCTTTCTCTCCAGGACGCCTGCACCCTCGT 9151

Qy        363 CACCCAACGCGCCACCCTCATGCAAACCATGCCCC---CCGGCACCATGACCACCCTCCA 419
              | ||   |||||  |  |||||||||  |  | || |    |||| |||||  ||||||||
Db       9152 CGCCGCCCGCGCAAAGCTCATGCAAGCGCTCCCACAAGGCGGCGCCATGGTCACCCTCCG 9211

Qy        420 CACCACCCCCCACCACATCACCCACCACCTCACCGCCCACGAAAACGACCTCGCCATCGC 479
              || ||    |  |  ||  | |||  |||    || ||||          | ||||
Db       9212 AGCCTCCGAGGAGGAAGTCCGCGACCTTCTCCAGCCCTACGATGGACGAGCTAGCCTCGC 9271

Qy        480 CGCCATCAACACCCCCACCTCCCTCGTCATCAGCGGCACCCCCCACACCGTCCAACACAT 539
              |||| ||||    ||   ||||  |||| ||   |||         || | ||      | ||
Db       9272 CGCCCTCAATGGGCCTCTCTCCACCGTCGTCGCTGGCGATGAAGACGCGGTGGTGGAGAT 9331

Qy        540 CACCACCCTCTGCCAACAACAAGGCATCAAAACCAAAACCCTCCCCACCAACCACGCCTT 599
              | ||  ||    | ||   |  ||    || |||| |  ||| | |  || ||||||||||
Db       9332 CGCCCGCCAGGCCGAAGCCCTCGGACGAAAGACCACACGCCTGCGCGTCAGCCACGCCTT 9391

Qy        600 CCACTCCCCCCACACCAACCCCATCCTCAACCAACTCCACCAGCACACCCAAACCCTCAC 659
              |||||| || ||||   ||   || ||| || |  ||| ||    | |||| | ||||||
Db       9392 CCACTCTCCGCACATGGACGGAATGCTCGACGACTTCCGCCGCGTCGCCCAGAGCCTCAC 9451

Qy        660 CTACCACCCACCCCACACCCCCCTCATCAC------------CGCCAACACCCCACCCGA 707
              |||||| ||  | | || |||| ||||| |            |||   | || |   | |
Db       9452 CTACCATCCCGCACGCATCCCCATCATCTCCAACGTCACCGGCGCGCGCGCCACGGACCA 9511

Qy        708 CCAACTCCTCACCCCCCACTACTGGACCCAACAAGCCCGCAACACCGTCGACTACGCCAC 767
              | | |||  | | |||  |||||||||  ||  || |  ||| |||||||||  || | |
Db       9512 CGAGCTCGCCTCGCCCGACTACTGGGTCCGCCACGTTCGCCACACCGTCCGCTTCCTCGA 9571

Qy        768 CACCACCCAAACCCTCCACCAACACGGCGTCACCACCTACATCGAACTCGGACCCGACAA 827
              |  |   |  |||| |||    | || |    |  ||   |||| ||||| ||  ||
Db       9572 CGGCGTACGTGCCCTTCACGCCGAAGGGGCACGCGTCTTTCTCGAGCTCGGGCCTCACGC 9631

Qy        828 CACCCTCACCACCCTCACCCACCACAACCTCCCCAACCCCCCCACCACCACCCTCACCCT 887
                |||| || ||||  | ||  ||  ||||     |    |     ||| | |     |
Db       9632 TGTCCTCTCCGCCCTTGCGCAAGACGCCCTCGGACAGGACGAAGGCACGTCGCCATGCGC 9691

Qy        888 CACCCACCCCCACCACCAC 906
              |  ||  |||  || || |
Db       9692 CTTCCTTCCCACCCTCCGC 9710


RESULT 11
US-08-858-003-1
; Sequence 1, Application US/08858003
; Patent No. 6060234
;  GENERAL INFORMATION:
;    APPLICANT:  Katz, Leonard
;    APPLICANT:  Stassi, Diane L.
;    APPLICANT:  Summers Jr., Richard G.
;    APPLICANT:  Ruan, Xiaoan
;    APPLICANT:  Pereda-Lopez, Ana
;    APPLICANT:  Kakavas, Stephan J.
;    TITLE OF INVENTION:  NOVEL POLYKETIDE DERIVATIVES
;    TITLE OF INVENTION:  AND RECOMBINANT METHODS FOR MAKING SAME
;    NUMBER OF SEQUENCES:  34
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Abbott Laboratories
;      STREET:  100 Abbott Park Rd.
;      CITY:  Abbott Park
;      STATE:  Illinois
;      COUNTRY:  USA
;      ZIP:  60064-3500
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/858,003
;      FILING DATE:  16-MAY-1979
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Dianne Casuto
;      REGISTRATION NUMBER:  P-40,943
;      REFERENCE/DOCKET NUMBER:  4952.US.P2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (847)-938-3137
;      TELEFAX:  (847)-938-2623
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  925 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
US-08-858-003-1

  Query Match            27.4%;  Score 253;  DB 3;  Length 925;
  Best Local Similarity  61.0%;  Pred. No. 8.1e-35;
  Matches  558;  Conservative    0;  Mismatches  255;  Indels  102;  Gaps    5;

Qy          4 GCATTCATCTGCTCCGGACAGGGCACCCAACGCCCCGGCATGGCCCACGGCCTCTACCAC 63
              ||  | || | | ||||||||||| |||||||||||||||||  |      | ||| |
Db         10 GCGGTGATGTTCACCGGACAGGGCTCCCAACGCCCCGGCATGGGACGACAGTTGTACGAG 69

Qy         64 ACCCACCCCGTCTTCGCCGCCGCACTCAACGACATCTGCACCCACCTCGACCCCCACCTC 123
                |  ||||||||||||   |||||  ||||  ||| | | | |  |  ||||   ||
Db         70 CACTTCCCCGTCTTCGCCCAGGCACTGGACGAGGTCTTCGCACTCGCCACCCCCGGACTA 129

Qy        124 GACCACCCCCTCCTCCCCCTCCTCACCCAAAACGACAACGACAACGAGGACGCGGCCGCA 183
              . | |      |  ||  || |                                 ||| ||||| |
Db        130 CGCGAGGTGATGTTCGACCCC--------------------------GACCAGGCCGAA 162

Qy        184 CTGCTCCAGCAGACCCGCTACGCCCAGCCCGCCCTCTTCGCCTTCCAGGTCGCCCTCCAC 243
              |||||  |  |||  | ||||||||  ||||||  |||||||||  |   |||||||  ||
Db        163 ACACTCCAACGCACCGACCACGCCCAGATCGCCCTGTTCGCCTTCGAAACCGCCCTCTAC 222
```

```
Qy        244 CGCCTCCTCACCGACGGCTACCACATCACCCCCCACTACTACGCCGGACACTCCCTCGGC 303
              ||    |||   ||   ||     | |    ||| ||     | |||||||||  ||||
Db        223 CG---ACTCTGGGAATCCTGGGGCCTGCGACCCGACATGGTCTGCGGACACTCGGTCGGA 279

Qy        304 GAAATCACCGCCGCCCACCTCGCCGGCATCCTCACCCTCACCGACGCCACCACCCTCATC 363
              |||||||||| |||||| || |||||| |||||||||| ||||||||  |  |||| ||
Db        280 GAAATCACCGCAGCCCACGTCTCCGGCACCCTCACCCTCCCCGACGCCGTCCACCTCGTC 339

Qy        364 ACCCAACGCGCCACCCTCATGCAAACCATGCCCCC---CGGCACCATGACCACCCTCCAC 420
              |||  ||||| |||||||||||||| | |||||||   |||| |||||  | || ||  |
Db        340 ACCACACGCGGCACCCTCATGCAAAACCTGCCCCCCGGCGGCGCCATGCTCGCCGTCGCC 399

Qy        421 ACCACCCCCCACCACATCACCCACCACCTCACCGCCCACGAAAACGACCTCGCCATCGCC 480
              |||  |||||||  | ||   | |||||||  |  |||| |  ||  | || ||||||||
Db        400 ACCGACCCCCACACCCTCCAACCCCACCTCGACAACCACCACGACACCATCTCCATCGCC 459

Qy        481 GCCATCAACACCCCCACCTCCCTCGTCATCAGCGGCACCCCCCACACCGTCCAACACATC 540
              |||||||||  ||||  | ||  |||| ||  ||||  || |  |||| |||| ||||||
Db        460 GCCATCAACGGCCCCCACGCCACCGTCCTCTCCGGCGACCGCACCACCCTCCACCACATC 519

Qy        541 ACCACCCTCTGCCAACAACAAGGCATCAAAACCAAAACCCTCCCCACCAACCACGCCTTC 600
              ||||||                |||| |||||| || || |||||| || |  |
Db        520 GCCACCCAAC--------------TCAACACCAAACCCTTCACCACCACCCTCAAC--- 561

Qy        601 CACTCCCCCCACACCAACCCCATCCTCAACCAACTCCACCAGCACACCCAAACCCTCACC 660
                                                                |||||||||
Db        562 ----------------------------------------------------ACCCTCACC 570

Qy        661 TACCACCCACCCCACACCCCCCTCATCACCGCCAACACCCCACCCGACCAACTCCTCACC 720
              |||||| || ||||| |||||||||| |     |||| |  |  |    |  || |  |
Db        571 CACCACCCCCCACACACACCCCTCATCAGCATGCTCACCGCCACACCCACCCACCCCGAC 630

Qy        721 CCCCACTACTGGACCCAACAAGCCCGCAACACCGTCGACTACGCCACCACCACCCAAACC 780
              ||  | |||||||||| ||   |  |    |||||  |||| ||  ||||  |||   |
Db        631 ACCACCCACTGGACCCAGCACATCACCGCACCCGTCCGCTACACCGACACCCTCCACCAC 690

Qy        781 CTCCACCAACACGGCGTCACCACCTACATCGAACTCGGACCCGACAACACCCTCACCACC 840
              |||||||| |||||| |||||||||||| |||||.|||| |||||||  ||||||||||  ||
Db        691 CTCCACCACCACGGCATCACCACCTACCTCGAAATCGGCCCCGACACCACCCTCACCGCC 750

Qy        841 CTCACCCACCACAACCTCCCCAACCCCCCCACCACCACCCTCACCCTCACCCACCCCCAC 900
              ||| ||| |  || |||||||| | || ||  |  || || |||| |  || |  ||||
Db        751 CTCGCCCGCACCACCCTCCCCACCACCACCCACCTCATCCCCACCACCCGCCGCAACCAC 810

Qy        901 CACCACCCCCAAACC 915
              || |   ||  | |
Db        811 AACGAAGTCCGCAGC 825
```

```
RESULT 12
US-09-078-166-1
; Sequence 1, Application US/09078166
; Patent No. 6063561
;  GENERAL INFORMATION:
;    APPLICANT:  Katz, Leonard
;    APPLICANT:  Stassi, Diane L.
;    APPLICANT:  Summers Jr., Richard G.
;    APPLICANT:  Ruan, Xiaoan
;    APPLICANT:  Pereda-Lopez, Ana
;    APPLICANT:  Kakavas, Stephan J.
;    TITLE OF INVENTION:  NOVEL POLYKETIDE DERIVATIVES
;    TITLE OF INVENTION:  AND RECOMBINANT METHODS FOR MAKING SAME
;    NUMBER OF SEQUENCES:  44
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Abbott Laboratories
;      STREET:  100 Abbott Park Rd.
;      CITY:  Abbott Park
;      STATE:  Illinois
;      COUNTRY:  USA
;      ZIP:  60064-3500
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/078,166
;      FILING DATE:  16-MAY-1979
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Dianne Casuto
;      REGISTRATION NUMBER:  P-40,943
;      REFERENCE/DOCKET NUMBER:  4952.US.P2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (847)-938-3137
;      TELEFAX:  (847)-938-2623
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  925 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
US-09-078-166-1

  Query Match             27.4%;  Score 253;  DB 3;  Length 925;
  Best Local Similarity   61.0%;  Pred. No. 8.1e-35;
  Matches  558;  Conservative    0;  Mismatches  255;  Indels  102;  Gaps    5;
```

```
Qy          4 GCATTCATCTGCTCCGGACAGGGCACCCAACGCCCCGGCATGGCCCACGGCCTCTACCAC 63
              ||  | || | | |||||||||| ||||||||||||||||||  |      | ||| |
Db         10 GCGGTGATGTTCACCGGACAGGGCTCCCAACGCCCCGGCATGGGACGACAGTTGTACGAG 69

Qy         64 ACCCACCCCGTCTTCGCCGCCGCACTCAACGACATCTGCACCCACCTCGACCCCCACCTC 123
                | ||||||||||||||   |||||  |||| |||  | | | |  |  ||||   ||
Db         70 CACTTCCCCGTCTTCGCCCAGGCACTGGACGAGGTCTTCGCACTCGCCACCCCCGGACTA 129

Qy        124 GACCACCCCCTCCTCCCCCTCCTCACCCAAAACGACAACGACAACGAGGACGCGGCCGCA 183
                |  |    |  ||  || |                  |||  |||||  |
Db        130 CGCGAGGTGATGTTCGACCCC----------------------------GACCAGGCCGAA 162

Qy        184 CTGCTCCAGCAGACCCGCTACGCCCAGCCCGCCCTCTTCGCCTTCCAGGTCGCCCTCCAC 243
                |||||  |  |||  | ||||||||  ||||||  ||||||||| |   |||||||  ||
Db        163 ACACTCCAACGCACCGACCACGCCCAGATCGCCCTGTTCGCCTTCGAAACCGCCCTCTAC 222

Qy        244 CGCCTCCTCACCGACGGCTACCACATCACCCCCCACTACTACGCCGGACACTCCCTCGGC 303
              ||    |||   ||   ||    |.|    ||| ||      | |||||||||  ||||
Db        223 CG---ACTCTGGGAATCCTGGGGCCTGCGACCCGACATGGTCTGCGGACACTCGGTCGGA 279

Qy        304 GAAATCACCGCCGCCCACCTCGCCGGCATCCTCACCCTCACCGACGCCACCACCCTCATC 363
              |||||||||| |||||| || |||||| |||||||||| ||||||||  |  |||| ||
Db        280 GAAATCACCGCAGCCCACGTCTCCGGCACCCTCACCCTCCCCGACGCCGTCCACCTCGTC 339

Qy        364 ACCCAACGCGCCACCCTCATGCAAACCATGCCCCC---CGGCACCATGACCACCCTCCAC 420
              |||  ||||| |||||||||||||| | |||||||   |||| |||||  | || ||  |
Db        340 ACCACACGCGGCACCCTCATGCAAAACCTGCCCCCCGGCGGCGCCATGCTCGCCGTCGCC 399

Qy        421 ACCACCCCCCACCACATCACCCACCACCTCACCGCCCACGAAAACGACCTCGCCATCGCC 480
              |||  |||||||  | ||   | |||||||  |  |||| |  ||  | || ||||||||
Db        400 ACCGACCCCCACACCCTCCAACCCCACCTCGACAACCACCACGACACCATCTCCATCGCC 459

Qy        481 GCCATCAACACCCCCACCTCCCTCGTCATCAGCGGCACCCCCCACACCGTCCAACACATC 540
              |||||||||  ||||  | ||  |||| ||  ||||  || |  |||| |||| ||||||
Db        460 GCCATCAACGGCCCCCACGCCACCGTCCTCTCCGGCGACCGCACCACCCTCCACCACATC 519

Qy        541 ACCACCCTCTGCCAACAACAAGGCATCAAAACCAAAACCCTCCCCACCAACCACGCCTTC 600
              ||||||                |||| |||||| || || |||||| || |  |
Db        520 GCCACCCAAC--------------TCAACACCAAACCCTTCACCACCACCCTCAAC--- 561

Qy        601 CACTCCCCCCACACCAACCCCATCCTCAACCAACTCCACCAGCACACCCAAACCCTCACC 660
                                                                |||||||||
Db        562 ----------------------------------------------------ACCCTCACC 570
```

```
Qy        661 TACCACCCACCCCACACCCCCCTCATCACCGCCAACACCCCACCCGACCAACTCCTCACC 720
              ||||||| || ||||| |||||||||| |     |||| |  |   |  | || |  |
Db        571 CACCACCCCCCACACACACCCCTCATCAGCATGCTCACCGCCACACCCACCCACCCCGAC 630

Qy        721 CCCCACTACTGGACCCAACAAGCCCGCAACACCGTCGACTACGCCACCACCACCCAAACC 780
              ||  | ||||||||| ||   |  |    ||||| |||| ||  |||| |||   |
Db        631 ACCACCCACTGGACCCAGCACATCACCGCACCCGTCCGCTACACCGACACCCTCCACCAC 690

Qy        781 CTCCACCAACACGGCGTCACCACCTACATCGAACTCGGACCCGACAACACCCTCACCACC 840
              |||||||| |||||| |||||||||| ||||| |||| ||||||| ||||||||| ||
Db        691 CTCCACCACCACGGCATCACCACCTACCTCGAAATCGGCCCCGACACCACCCTCACCGCC 750

Qy        841 CTCACCCACCACAACCTCCCCAACCCCCCCACCACCACCCTCACCCTCACCCACCCCCAC 900
              ||| ||| |  || |||||||| | || ||  |  || || |||| |  || |  ||||
Db        751 CTCGCCCGCACCACCCTCCCCACCACCACCCACCTCATCCCCACCACCCGCCGCAACCAC 810

Qy        901 CACCACCCCCAAACC 915
              || |   ||  | |
Db        811 AACGAAGTCCGCAGC 825
```

RESULT 13
US-08-997-467-1
; Sequence 1, Application US/08997467
; Patent No. 6200813
; GENERAL INFORMATION:
; APPLICANT: Katz, Leonard
; APPLICANT: Stassi, Diane L.
; APPLICANT: Summers Jr., Richard G.
; APPLICANT: Ruan, Xiaoan
; APPLICANT: Pereda-Lopez, Ana
; APPLICANT: Kakavas, Stephan J.
; TITLE OF INVENTION: NOVEL POLYKETIDE DERIVATIVES
; TITLE OF INVENTION: AND RECOMBINANT METHODS FOR MAKING SAME
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Abbott Laboratories
; STREET: 100 Abbott Park Rd.
; CITY: Abbott Park
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60064-3500
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/997,467
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/858,003
; FILING DATE: 16-MAY-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Dianne Casuto
; REGISTRATION NUMBER: P-40,943
; REFERENCE/DOCKET NUMBER: 4952.US.P2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (847)-938-3137
; TELEFAX: (847)-938-2623
; TELEX:
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 925 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
US-08-997-467-1

Query Match 27.4%; Score 253; DB 4; Length 925;
Best Local Similarity 61.0%; Pred. No. 8.1e-35;

Matches 558; Conservative 0; Mismatches 255; Indels 102; Gaps 5;

```
Qy          4 GCATTCATCTGCTCCGGACAGGGCACCCAACGCCCCGGCATGGCCCACGGCCTCTACCAC 63
              ||  | || | | |||||||||| ||||||||||||||||||  |      | ||| |
Db         10 GCGGTGATGTTCACCGGACAGGGCTCCCAACGCCCCGGCATGGGACGACAGTTGTACGAG 69

Qy         64 ACCCACCCCGTCTTCGCCGCCGCACTCAACGACATCTGCACCCACCTCGACCCCCACCTC 123
                | |||||||||||||  |||||  |||| ||| | | | |  |  ||||   ||
Db         70 CACTTCCCCGTCTTCGCCCAGGCACTGGACGAGGTCTTCGCACTCGCCACCCCCGGACTA 129

Qy        124 GACCACCCCCTCCTCCCCCTCCTCACCCAAAACGACAACGACAACGAGGACGCGGCCGCA 183
              | |    | ||  || |                                ||| |||| |
Db        130 CGCGAGGTGATGTTCGACCCC-------------------------GACCAGGCCGAA 162

Qy        184 CTGCTCCAGCAGACCCGCTACGCCCAGCCCGCCCTCTTCGCCTTCCAGGTCGCCCTCCAC 243
              |||| |  |||  | ||||||||  |||||| |||||||| |   |||||||  ||
Db        163 ACACTCCAACGCACCGACCACGCCCAGATCGCCCTGTTCGCCTTCGAAACCGCCCTCTAC 222

Qy        244 CGCCTCCTCACCGACGGCTACCACATCACCCCCCACTACTACGCCGGACACTCCCTCGGC 303
              ||    |||   ||   ||     | |    ||| ||  .  |  |||||||||  ||||
Db        223 CG---ACTCTGGGAATCCTGGGGCCTGCGACCCGACATGGTCTGCGGACACTCGGTCGGA 279

Qy        304 GAAATCACCGCCGCCCACCTCGCCGGCATCCTCACCCTCACCGACGCCACCACCCTCATC 363
              |||||||||| |||||| || |||||| |||||||| ||||||||  |  |||| ||
Db        280 GAAATCACCGCAGCCCACGTCTCCGGCACCCTCACCCTCCCCGACGCCGTCCACCTCGTC 339

Qy        364 ACCCAACGCGCCACCCTCATGCAAACCATGCCCCC---CGGCACCATGACCACCCTCCAC 420
              |||  ||||| |||||||||||||| | |||||||   |||| |||||  | || ||  |
Db        340 ACCACACGCGGCACCCTCATGCAAAACCTGCCCCCCGGCGGCGCCATGCTCGCCGTCGCC 399

Qy        421 ACCACCCCCCACCACATCACCCACCACCTCACCGCCCACGAAAACGACCTCGCCATCGCC 480
              |||  |||||||  | ||  | |||||||  |  |||| |  ||  |  || |||||||||
Db        400 ACCGACCCCCCACACCCTCCAACCCCACCTCGACAACCACCACGACACCATCTCCATCGCC 459

Qy        481 GCCATCAACACCCCCACCTCCCTCGTCATCAGCGGCACCCCCCCACACCGTCCAACACATC 540
              |||||||||  ||||  | ||  |||| ||  ||||  || |  |||| |||| ||||||
Db        460 GCCATCAACGGCCCCCACGCCACCGTCCTCTCCGGCGACCGCACCACCCTCCACCACATC 519

Qy        541 ACCACCCTCTGCCAACAACAAGGCATCAAAACCAAAAACCCTCCCCACCAACCACGCCTTC 600
              ||||||                 |||| |||||| || || |||||| || |  |
Db        520 GCCACCCAAC--------------TCAACACCAAACCCTTCACCACCACCCTCAAC--- 561

Qy        601 CACTCCCCCCACACCAACCCCATCCTCAACCAACTCCACCAGCACACCCAAACCCTCACC 660
                                                                |||||||||
Db        562 ----------------------------------------------------ACCCTCACC 570

Qy        661 TACCACCCACCCCACACCCCCCTCATCACCGCCAACACCCCACCCGACCAACTCCTCACC 720
              ||||||| || ||||| |||||||||| |     |||| |  |   |  | || |  |
Db        571 CACCACCCCCCACACACACCCCTCATCAGCATGCTCACCGCCACACCCACCCACCCCGAC 630

Qy        721 CCCCACTACTGGACCCAACAAGCCCGCAACACCGTCGACTACGCCACCACCACCCAAACC 780
              ||  | ||||||||| ||   |  |    ||||| |||| ||  |||| |||   |
Db        631 ACCACCCACTGGACCCAGCACATCACCGCACCCGTCCGCTACACCGACACCCTCCACCAC 690

Qy        781 CTCCACCAACACGGCGTCACCACCTACATCGAACTCGGACCCGACAACACCCTCACCACC 840
              |||||||| |||||| |||||||||| ||||| |||| ||||||| ||||||||| ||
Db        691 CTCCACCACCACGGCATCACCACCTACCTCGAAATCGGCCCCGACACCACCCTCACCGCC 750

Qy        841 CTCACCCACCACAACCTCCCCAACCCCCCCACCACCACCCTCACCCTCACCCACCCCCAC 900
              ||| ||| |  || |||||||| | || ||  |  || || |||| |  || |  ||||
Db        751 CTCGCCCGCACCACCCTCCCCACCACCACCCACCTCATCCCCACCACCCGCCGCAACCAC 810

Qy        901 CACCACCCCCAAACC 915
              || |   ||  | |
Db        811 AACGAAGTCCGCAGC 825
```

RESULT 14
US-09-105-537-1
; Sequence 1, Application US/09105537A
; Patent No. 6265202
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.

```
;  APPLICANT: Liu, H.
;  APPLICANT: Xue, Y.
;  APPLICANT: Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/105,537A
;  CURRENT FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 1
;   LENGTH: 15872
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-105-537-1

  Query Match           25.9%;  Score 239.4;  DB 4;  Length 15872;
  Best Local Similarity  58.0%;  Pred. No. 1.9e-32;
  Matches  523;  Conservative    0;  Mismatches  336;  Indels   42;  Gaps    4;

Qy          1 ACCGCATTCATCTGCTCCGGACAGGGCACCCAACGCCCCGGCATGGCCCACGGCCTCTAC 60
              || || ||| | | | |||| |||||  | || |||  || |||||| | | |  ||   |
Db       4754 ACGGCCTTCCTGTTCACCGGCCAGGGAGCGCAGCGCGTCGCCATGGGCGAGGAACTGCGC 4813

Qy         61 CACACCCACCCCGTCTTCGCCGCCGCACTCAACGACATCTGCACCCACCTCGACCCCCAC 120
                | | |||||||| |||||||||| ||| ||  | | | | |   ||||||||  |||
Db       4814 GCCGCGCACCCCGTGTTCGCCGCCGCCCTCGACACCGTGTACGCGGCCCTCGACCGTCAC 4873

Qy        121 CTCGACCACCCCCTCCTCCCCCTCCTCACCCAAAACGACAACGACAACGAGGACGCGGCC 180
              |||||||  || || |                          || | ||  | ||| |
Db       4874 CTCGACCGGCCGCTGCGG----------------------GAGATCGTCGCCGCCGGG 4909

Qy        181 GCACTGCTCCAGCAGACCCGCTACGCCCAGCCCGCCCTCTTCGCCTTCCAGGTCGCCCTC 240
              |    |||  | |  |||   ||| ||||||||||||||||||||| |||| || ||
Db       4910 GAGGAGCTGGACCTCACCGCGTACACCCAGCCCGCCCTCTTCGCCTTCGAGGTGGCGCTG 4969

Qy        241 CACCGCCTCCTCACCGACGGCTACCACATCACCCCCCACTACTACGCCGGACACTCCCTC 300
              |||||||||   |||  | ||  | ||  |||| ||      | |||| ||||||  ||
Db       4970 TTCCGCCTCCT---CGAACACCACGGCCTCGTCCCCGACCTGCTCACCGGCCACTCCGTC 5026

Qy        301 GGCGAAATCACCGCCGCCCACCTCGCCGGCATCCTCACCCTCACCGACGCCACCACCCTC 360
              ||||| ||| ||||||| ||| |||||||  |||||  |||||  |||||||  |    |||
Db       5027 GGCGAGATCGCCGCCGCGCACGTCGCCGGTGTCCTCTCCCTCGACGACGCCGCACGTCTC 5086

Qy        361 ATCACCCAACGCGCCACCCTCATGCAAACCATGCCC---CCCGGCACCATGACCACCCTC 417
              |||||   ||||  |   ||||||||  |     | |     ||||  | ||||  | || |
Db       5087 GTCACCGCCCGCGGCCGGCTCATGCAGTCGGCCCGCGAGGGCGGCGCGATGATCGCCGTG 5146

Qy        418 CACACCACCCCCCACCACATCACCCACCACCTCACCGCCCACGAAAACGACCTCGCCATC 477
              ||  |  |      |  ||  | |   ||| |  | | |||||    |     ||||| ||
Db       5147 CAGGCGGGCGAGGCCGAGGTCGTCGAGTCCCTGAAGGGCTACGAGGGCAGGGTCGCCGTC 5206

Qy        478 GCCGCCATCAACACCCCCACCTCCCTCGTCATCAGCGGCACCCCCCACACCGTCCAACAC 537
              |||||| |||||   |||||| || | ||| ||  ||||   |  |  || ||| | |  |
Db       5207 GCCGCCGTCAACGGACCCACCGCCGTGGTCGTCTCCGGCGACGCGGACGCCGCCGAGGAG 5266

Qy        538 ATCACCACCCTCTGCCAACAACAAGGCATCAAAACCAAAACCCTCCCCACCAACCACGCC 597
              |||  | ||| | ||      ||  |||       |||   | || | |  || |||||||
Db       5267 ATCCGCGCCGTATGGGCGGGACGCGGCCGGCGCACCCGCAGGCTGCGCGTCAGCCACGCC 5326

Qy        598 TTCCACTCCCCCCACACCAACCCCATCCTCAACCAACTCCACCAGCACACCCAAACCCTC 657
              |||||||||| ||||   ||  | ||||| || |  ||| || |  | || |   |||
Db       5327 TTCCACTCCCCGCACATGGACGACGTCCTCGACGAGTTCCTCCGGGTCGCCGAGGGCCTG 5386

Qy        658 ACCTACCACCCACCCCACACCCCCCTCATCACCGCCAACAC----------CCCACCC 705
              |||| | |   || |  | |||| || || || |   |||          | |  ||
Db       5387 ACCTTCGAGGAGCCGCGGATCCCCGTCGTCTCCACGGTCACCGGCGCGCTCGTCACGTCC 5446

Qy        706 GACCAACTCCTCACCCCCCACTACTGGACCCAACAAGCCCGCAACACCGTCGACTACGCC 765
              | | | |||  | | |||   ||||||  | | | ||   |||      ||||   || |
Db       5447 GGCGAGCTCACCTCGCCCGCGTACTGGGTCGACCAGATCCGGCGGCCCGTGCGCTTCCTG 5506

Qy        766 ACCACCACCCAAACCCTCCACCAACACGGCGTCACCACCTACATCGAACTCGGACCCGAC 825
                | ||  ||  |||||   |   || | ||  |||  |  | ||||  |||| ||||||

Db       5507 GACGCCGTCCGCACCCTGGCCGCCCAGGACGCGACCGTCCTCGTCGAGATCGGCCCCGAC 5566

Qy        826 AACACCCTCACCACCCTCACCCACCACAACCTCCCCAACCCCCCCCACCACCACCCTCACC 885
                |  ||||||  | ||| || |  |    ||| |   | | |   | ||| |   | ||
Db       5567 GCCGTCCTCACGGCACTCGCCGAGGAGGCTCTCGCGCCCGGCACGGACGCCCCGGACGCC 5626

Qy        886 C 886
              |
Db       5627 C 5627


RESULT 15
US-09-428-517-1
; Sequence 1, Application US/09428517
; Patent No. 6251636
; GENERAL INFORMATION:
;  APPLICANT: Betlach, Mary C.
;  APPLICANT: Shah, Sanjay Krishnakant
;  APPLICANT: McDaniel, Robert
;  APPLICANT: Tang, Li
;  TITLE OF INVENTION: RECOMBINANT OLEANDOLIDE POLYKETIDE SYNTHASE
;  FILE REFERENCE: 30062-20029.00
;  CURRENT APPLICATION NUMBER: US/09/428,517
;  CURRENT FILING DATE: 1999-10-28
;  EARLIER APPLICATION NUMBER: 60/120,254
;  EARLIER FILING DATE: 1999-02-16
;  EARLIER APPLICATION NUMBER: 60/106,100
;  EARLIER FILING DATE: 1998-10-29
;  NUMBER OF SEQ ID NOS: 12
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
;   LENGTH: 50937
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artifical Sequence: Recombinant DNA
US-09-428-517-1

  Query Match           25.2%;  Score 232.6;  DB 4;  Length 50937;
  Best Local Similarity  57.8%;  Pred. No. 2.8e-31;
  Matches  528;  Conservative    0;  Mismatches  349;  Indels   36;  Gaps    5;

Qy          3 CGCATTCATCTGCTCCGGACAGGGCACCCAACGCCCCGGCATGGCCCACGGCCTCTACCA 62
              ||| ||| ||| |||||| ||||||| ||| |  |||||||||| | |  | ||| ||||
Db       7457 CGCCTTCCTCTTCTCCGGGCAGGGCAGCCAGCAGCCCGGCATGGGCAAGCGACTCCACCA 7516

Qy         63 CACCCACCCCGTCTTCGCCGCCGCACTCAACGACATCTGCACCCACCTCGACCCCCACCT 122
                ||||| ||||   | ||| ||  ||||  ||||| || | |||||| |||||||
Db       7517 GGTGTTCCCCGGCTTCCGGGACGCCCTGGACGAGGTCTGCGCCGAACTCGACACCCACCT 7576

Qy        123 CGACCACCCCCTCCTCCCCCTCCTCACCCAAAACGACAACGACAACGAGGACGCGG---- 178
              || ||  | ||||  ||||      |  ||    |      |||       |  ||| |
Db       7577 CGGCCGACTCCTCGGCCCCGAGGCCGGCCCGCCCCTGCGCGACGTGATGTTCGCCGAGCG 7636

Qy        179 --------------CCGCACTGCTCCAGCAGACCCGCTACGCCCAGCCCGCCCTCTTCGC 224
                           ||| ||||||    |||||| |||| ||||| ||||||||||||
Db       7637 GGGCACGGCGCACAGCGCCCTGCTCTCCGAGACCCACTACACCCAGGCCGCCCTCTTCGC 7696

Qy        225 CTTCCAGGTCGCCCTCCACCGCCTCCTCACCGACGGCTACCACATCACCCCCCACTACTA 284
              | | |  |||||||  |||||||||   | |  |   ||  | |  ||| || ||
Db       7697 CCTGGAAACCGCCCTCTTCCGCCTCCTGGTCCAGTGGGGCC---TGAAACCCGACCACCT 7753

Qy        285 CGCCGGACACTCCCTCGGCGAAATCACCGCCGCCCACCTCGCCGGCATCCTCACCCTCAC 344
              ||| || ||||||  |||||||  ||| | |||||||||    || |||||||||  |||  |
Db       7754 CGCAGGCCACTCCGTCGGCGAGATCGCGGCCGCCCACGCAGCAGGCATCCTCGACCTGTC 7813

Qy        345 CGACGCCACCACCCTCATCACCCAACGCGCCACCCTCATGCAAACCATGCC---CCCCGG 401
              ||||||  ||   |||  |  ||   |||| | |  | ||||   || ||||   | |||
Db       7814 CGACGCGGCCGAACTCGTGGCCACCCGCGGCGCGTTGATGCGTTCCCTGCCCGGCGGCGG 7873

Qy        402 CACCATGACCACCCTCCACACCACCCCCCACCACATCACCCACCACCTCACCGCCCACGA 461
              |  ||||  | |  ||||  |  ||      | |  || | | ||  ||   || ||  ||
Db       7874 CGTCATGCTCTCGGTCCAGGCACCCGAGTCCGAGGTCGCACCCCTGCTGCTCGGCCGTGA 7933
```

```
Qy       462 AAACGACCTCGCCATCGCCGCCATCAACACCCCCACCTCCCTCGTCATCAGCGGCACCCC 521
             |  ||  |||  |  |  ||||||  |.|||  ||||  |  |  |  |||  |    ||||   |
Db      7934 GGCCCACGTCGGCCTGGCCGCCGTGAACGGCCCCGACGCGGTGGTCGTGTCCGGCGAGCG 7993

Qy       522 CCACACCGTCCAACACATCACCACCCTCTGCCAACAACAAGGCATCAAAACCAAAACCCT 581
             |  |  ||||     ||||        ||  ||    |    |||  ||||| |     |||
Db      7994 CGGCCACGTCGCCGCCATCGAACAGATCCTCCGGGACAGGGGCCGCAAAAGCCGGTACCT 8053

Qy       582 CCCCACCAACCACGCCTTCCACTCCCCCCACACCAACCCCATCCTCAACCAACTCCACCA 641
             | |  || ||||||||||||||||||| | ||    | ||  | ||  |  |  ||  | |
Db      8054 GCGCGTCAGCCACGCCTTCCACTCCCCGCTCATGGAACCGGTGCTGGAGGAGTTCGCCGA 8113

Qy       642 GCACACCCAAACCCTCACCTACCACCCACCCCACACCCCCCTCATCACCGCCAACACC-- 699
             |  |      ||| |||| ||   ||||   ||| ||||||  ||  ||   |  ||||
Db      8114 AGCCGTCGCCGGCCTGACCTTCCGGGCACCGACCACACCCCTCGTCTCCAACCTCACCGG 8173

Qy       700 ----CCACCCGACCAAC------TCCTCACCCCCCACTACTGGACCCAACAAGCCCGCAA 749
                 |||  |||| | |       |   ||| |||  |||||||  ||  || | |||  |
Db      8174 CGCACCAGTCGACGACCGGACCATGGCCACGCCCGCCTACTGGGTCCGGCACGTCCGGGA 8233

Qy       750 CACCGTCGACTACGCCACCACCACCCAAACCCTCCACCAACACGGCGTCACCACCTACAT 809
             | |||  || || |  |  || ||   | |||    |||  |||  |  || || | |
Db      8234 AGCGGTCCGCTTCGGCGACGGCATCCGGGCACTCGGGAAACTGGGCACCGGCAGCTTCCT 8293

Qy       810 CGAACTCGGACCCGACAACACCCTCACCACCCTCACCCACCACAACCTCCCCAACCCCCC 869
              ||| |||| || |||  |  |||||| || |  | | |    | || ||  | ||||
Db      8294 GGAAGTCGGGCCGGACGGCGTCCTCACCGCCATGGCGCGCGCATGCGTCACCGCCGCCCC 8353

Qy       870 CACCACCACCCTC 882
                ||  || |
Db      8354 GGAGCCCGGCCAC 8366
```

Search completed: June 18, 2003, 01:07:32
Job time : 50.0267 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 17, 2003, 17:56:34 ; Search time 109.197 Seconds
(without alignments)
12253.192 Million cell updates/sec

Title: US-09-914-286-1_COPY_7906_8829
Perfect score: 924
Sequence: 1 accgcattcatctgctccgg.....accccaaacccacctcctc 924

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1029858 seqs, 724030393 residues

Total number of hits satisfying chosen parameters: 2059716

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
11: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
12: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
13: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
14: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
c	1	256.8	27.8	12733	9	US-10-032-393-47	Sequence 47, Appl
c	2	256.8	27.8	12739	9	US-10-032-393-8	Sequence 8, Appli
	3	253	27.4	925	9	US-09-735-056-1	Sequence 1, Appli
	4	239.4	25.9	15872	9	US-09-860-846-1	Sequence 1, Appli
	5	239.4	25.9	15872	9	US-09-988-384B-1	Sequence 1, Appli
	6	239.4	25.9	15872	9	US-09-836-821-1	Sequence 1, Appli
	7	239.4	25.9	15872	10	US-09-861-289-1	Sequence 1, Appli
	8	232.6	25.2	50937	9	US-09-808-880-1	Sequence 1, Appli
	9	220.2	23.8	13842	9	US-09-860-846-30	Sequence 30, Appl
	10	220.2	23.8	13842	9	US-09-988-384B-30	Sequence 30, Appl
	11	220.2	23.8	13842	9	US-09-836-821-30	Sequence 30, Appl
	12	220.2	23.8	13842	10	US-09-861-289-30	Sequence 30, Appl
	13	220.2	23.8	36778	9	US-09-860-846-5	Sequence 5, Appli
	14	220.2	23.8	36778	9	US-09-836-821-5	Sequence 5, Appli
	15	220.2	23.8	36778	10	US-09-861-289-5	Sequence 5, Appli
	16	220.2	23.8	37948	9	US-09-988-384B-5	Sequence 5, Appli
	17	220.2	23.8	38506	9	US-09-793-708-19	Sequence 19, Appl
	18	218.8	23.7	1030	9	US-09-735-056-2	Sequence 2, Appli
c	19	211.2	22.9	1075	10	US-09-864-761-19241	Sequence 19241, A
c	20	211.2	22.9	1403	10	US-09-864-761-2513	Sequence 2513, Ap
	21	209.8	22.7	1635	10	US-09-864-761-20241	Sequence 20241, A
	22	209.8	22.7	1973	10	US-09-864-761-3471	Sequence 3471, Ap
	23	192.2	20.8	4041	9	US-09-860-846-36	Sequence 36, Appl
	24	192.2	20.8	4041	9	US-09-988-384B-36	Sequence 36, Appl
	25	192.2	20.8	4041	9	US-09-836-821-36	Sequence 36, Appl
	26	192.2	20.8	4041	10	US-09-861-289-36	Sequence 36, Appl
	27	191.6	20.7	15720	9	US-10-025-380-1058	Sequence 1058, Ap
	28	191.6	20.7	15720	10	US-09-922-217-1058	Sequence 1058, Ap
	29	191.6	20.7	15720	10	US-09-833-263-1058	Sequence 1058, Ap
	30	176.6	19.1	4689	9	US-09-860-846-34	Sequence 34, Appl
	31	176.6	19.1	4689	9	US-09-988-384B-34	Sequence 34, Appl
	32	176.6	19.1	4689	9	US-09-836-821-34	Sequence 34, Appl
	33	176.6	19.1	4689	10	US-09-861-289-34	Sequence 34, Appl
	34	162.2	17.6	11220	9	US-09-860-846-32	Sequence 32, Appl
	35	162.2	17.6	11220	9	US-09-988-384B-32	Sequence 32, Appl
	36	162.2	17.6	11220	9	US-09-836-821-32	Sequence 32, Appl
	37	162.2	17.6	11220	10	US-09-861-289-32	Sequence 32, Appl
	38	152.6	16.5	1064	10	US-09-804-682-29	Sequence 29, Appl
	39	146.2	15.8	1065	10	US-09-804-682-33	Sequence 33, Appl
c	40	131.8	14.3	155074	9	US-10-026-188-6	Sequence 6, Appli
c	41	130	14.1	931	9	US-10-198-846-8585	Sequence 8585, Ap
c	42	127.2	13.8	1981	10	US-09-864-761-4745	Sequence 4745, Ap
c	43	124.6	13.5	887	10	US-09-864-761-21484	Sequence 21484, A
	44	124.2	13.4	1278	9	US-10-124-800-9	Sequence 9, Appli
	45	124.2	13.4	8730	9	US-10-124-800-1	Sequence 1, Appli

ALIGNMENTS

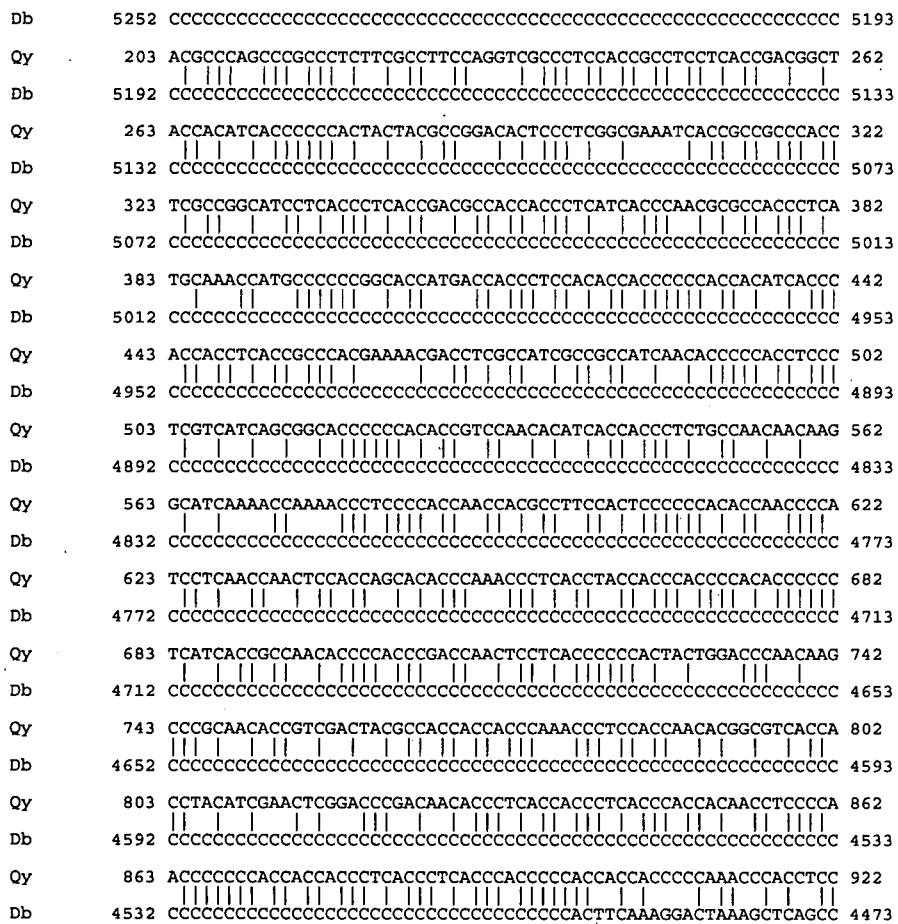
RESULT 1
US-10-032-393-47/c
; Sequence 47, Application US/10032393
; Publication No. US20030027286A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Wall, Daniel
; APPLICANT: Gross, Molly
; TITLE OF INVENTION: BACTERIAL PROMOTERS AND METHODS OF USE
; FILE REFERENCE: ELITRA.010A
; CURRENT APPLICATION NUMBER: US/10/032,393
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: 60/259,434
; PRIOR FILING DATE: 2000-12-27
; PRIOR APPLICATION NUMBER: 09/948,993
; PRIOR FILING DATE: 2001-09-06
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; NUMBER OF SEQ ID NOS: 68
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 47
; LENGTH: 12733
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Vector pEPEF14
US-10-032-393-47

Query Match 27.8%; Score 256.8; DB 9; Length 12733;
Best Local Similarity 55.3%; Pred. No. 2.6e-46;
Matches 498; Conservative 0; Mismatches 402; Indels 0; Gaps 0;

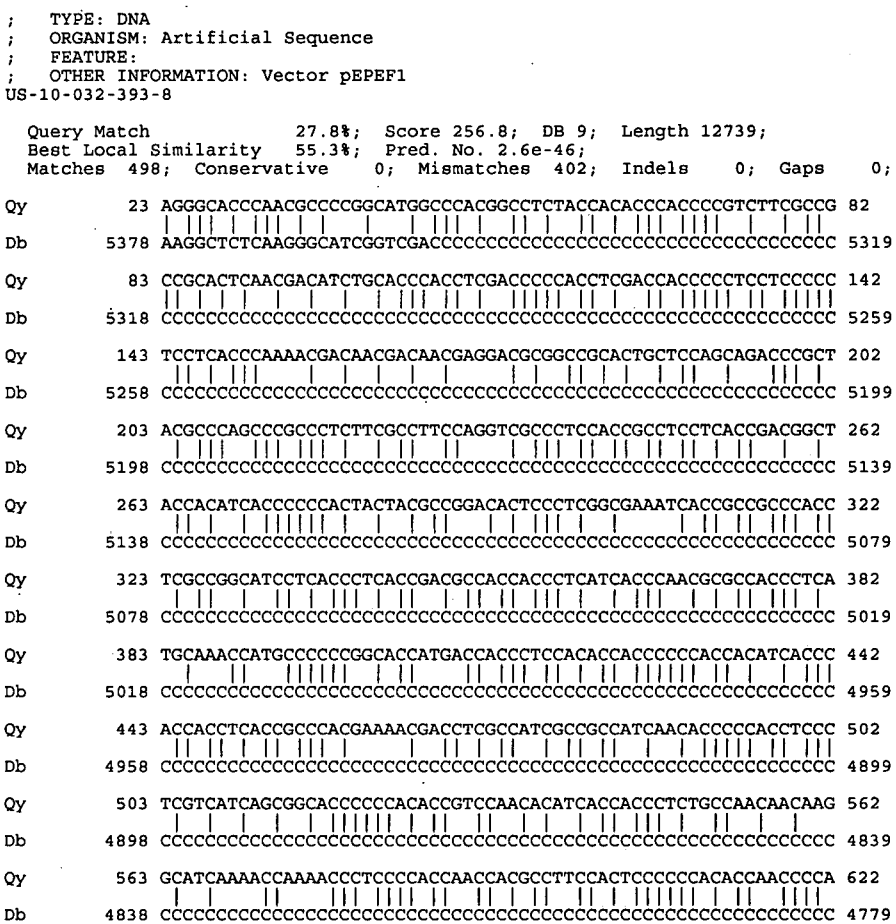
```
Qy         23 AGGGCACCCAACGCCCCGGCATGGCCCACGGCCTCTACCACACCCACCCCGTCTTCGCCG 82
              ||| | ||| | |   |   | ||| |  || |  || | ||| |||| |  | ||
Db       5372 AAGGCTCTCAAGGGCATCGGTCGACCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 5313

Qy         83 CCGCACTCAACGACATCTGCACCCACCTCGACCCCCACCTCGACCACCCCCTCCTCCCCC 142
              || | | |  | | |  | ||| || |  ||||| || |  || ||||| || |||||
Db       5312 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 5253

Qy        143 TCCTCACCCAAAACGACAACGACAACGAGGACGCGGCCGCACTGCTCCAGCAGACCCGCT 202
                || | |||    |  |  |  |  |     | |  || | |  | ||  |   ||| |
```



RESULT 2
US-10-032-393-8/c
; Sequence 8, Application US/10032393
; Publication No. US20030027286A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Wall, Daniel
; APPLICANT: Gross, Molly
; TITLE OF INVENTION: BACTERIAL PROMOTERS AND METHODS OF USE
; FILE REFERENCE: ELITRA.010A
; CURRENT APPLICATION NUMBER: US/10/032,393
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: 60/259,434
; PRIOR FILING DATE: 2000-12-27
; PRIOR APPLICATION NUMBER: 09/948,993
; PRIOR FILING DATE: 2001-09-06
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; NUMBER OF SEQ ID NOS: 68
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 12739



RESULT 3
US-09-735-056-1

```
; Sequence 1, Application US/09735056
; Publication No. US20030013662A1
;  GENERAL INFORMATION:
;    APPLICANT:  Katz, Leonard
;    APPLICANT:  Stassi, Diane L.
;    APPLICANT:  Summers Jr., Richard G.
;    APPLICANT:  Ruan, Xiaoan
;    APPLICANT:  Pereda-Lopez, Ana
;    APPLICANT:  Kakavas, Stephan J.
;    TITLE OF INVENTION:  NOVEL POLYKETIDE DERIVATIVES
;    TITLE OF INVENTION:  AND RECOMBINANT METHODS FOR MAKING SAME
;    NUMBER OF SEQUENCES:  34
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Abbott Laboratories
;      STREET:  100 Abbott Park Rd.
;      CITY:  Abbott Park
;      STATE:  Illinois
;      COUNTRY:  USA
;      ZIP:  60064-3500
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/735,056
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/858,003
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Dianne Casuto
;      REGISTRATION NUMBER:  P-40,943
;      REFERENCE/DOCKET NUMBER:  4952.US.P2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (847)-938-3137
;      TELEFAX:  (847)-938-2623
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  925 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
US-09-735-056-1

  Query Match             27.4%;  Score 253;  DB 9;  Length 925;
  Best Local Similarity   61.0%;  Pred. No. 1.8e-45;
  Matches  558;  Conservative    0;  Mismatches  255;  Indels  102;  Gaps    5;

Qy          4 GCATTCATCTGCTCCGGACAGGGCACCCAACGCCCCGGCATGGCCCACGGCCTCTACCAC 63
              ||  | || | | |||||||||||| ||||||||||||||||||  |      | ||| |
Db         10 GCGGTGATGTTCACCGGACAGGGCTCCCAACGCCCCGGCATGGGACGACAGTTGTACGAG 69

Qy         64 ACCCACCCCGTCTTCGCCGCCGCACTCAACGACATCTGCACCCACCTCGACCCCCACCTC 123
                |  |||||||||||||   |||||  ||||  ||| | | | |  |  ||||   ||
Db         70 CACTTCCCCGTCTTCGCCCAGGCACTGGACGAGGTCTTCGCACTCGCCACCCCCGGACTA 129

Qy        124 GACCACCCCCTCCTCCCCCTCCTCACCCAAAACGACAACGACAACGAGGACGCGGCCGCA 183
                | |     |  ||  || |                                 |||  ||||| |
Db        130 CGCGAGGTGATGTTCGACCCC-------------------------GACCAGGCCGAA 162

Qy        184 CTGCTCCAGCAGACCCGCTACGCCCAGCCCGCCCTCTTCGCCTTCCAGGTCGCCCTCCAC 243
                |||| |  |||  | ||||||||  ||||||  |||||||||  |   |||||||  ||
Db        163 ACACTCCAACGCACCGACCACGCCCAGATCGCCCTGTTCGCCTTCGAAACCGCCCTCTAC 222

Qy        244 CGCCTCCTCACCGACGGCTACCACATCACCCCCCACTACTACGCCGGACACTCCCTCGGC 303
                ||    |||   ||   ||     | |    ||| ||     |  |||||||||  ||||
Db        223 CG---ACTCTGGGAATCCTGGGGCCTGCGACCCGACATGGTCTGCGGACACTCGGTCGGA 279

Qy        304 GAAATCACCGCCGCCCACCTCGCCGGCATCCTCACCCTCACCGACGCCACCACCCTCATC 363
              |||||||||||| |||||| || |||||| |||||||||| ||||||||  |  |||| ||

Db        280 GAAATCACCGCAGCCCACGTCTCCGGCACCCTCACCCTCCCCGACGCCGTCCACCTCGTC 339

Qy        364 ACCCAACGCGCCACCCTCATGCAAACCATGCCCCC---CGGCACCATGACCACCCTCCAC 420
              |||  |||||  |||||||||||||||| | |||||||    |||| |||||  | || ||  |
Db        340 ACCACACGCGGCACCCTCATGCAAAACCTGCCCCCCGGCGGCGCCATGCTCGCCGTCGCC 399

Qy        421 ACCACCCCCCCACCACATCACCCACCACCTCACCGCCCACGAAAACGACCTCGCCATCGCC 480
              |||  |||||||  | ||   | |||||||  |  ||||  |  ||  |  || ||||||||
Db        400 ACCGACCCCCACACCCTCCAACCCCACCTCGACAACCACCACGACACCATCTCCATCGCC 459

Qy        481 GCCATCAACACCCCCACCTCCCTCGTCATCAGCGGCACCCCCCACACCGTCCAACACATC 540
              |||||||||  ||||  | ||  ||||  ||  ||||  ||  |  ||||  |||| ||||||
Db        460 GCCATCAACGGCCCCCACGCCACCGTCCTCTCCGGCGACCGCACCACCCTCCACCACATC 519

Qy        541 ACCACCCTCTGCCAACAACAAGGCATCAAAACCAAAAACCCTCCCCACCAACCACGCCTTC 600
              ||||||                    |||| ||||||  ||  ||  ||||||  ||  |  |
Db        520 GCCACCCAAC--------------TCAACACCAAACCCTTCACCACCACCCTCAAC--- 561

Qy        601 CACTCCCCCCACACCAACCCCATCCTCAACCAACTCCACCAGCACACCCAAACCCTCACC 660
                                                                  |||||||||
Db        562 ----------------------------------------------------ACCCTCACC 570

Qy        661 TACCACCCACCCCACACCCCCCTCATCACCGCCAACACCCCACCCGACCAACTCCTCACC 720
              |||||||  || ||||| ||||||||||  |      |||| |  |    |  | || |  |
Db        571 CACCACCCCCCACACACACCCCTCATCAGCATGCTCACCGCCACACCCACCCACCCCGAC 630

Qy        721 CCCCACTACTGGACCCAACAAGCCCGCAACACCGTCGACTACGCCACCACCACCCAAACC 780
              ||  | ||||||||||| ||    |  |     |||||  |||| ||  ||||  |||   |
Db        631 ACCACCCACTGGACCCAGCACATCACCGCACCCGTCCGCTACACCGACACCCTCCACCAC 690

Qy        781 CTCCACCAACACGGCGTCACCACCTACATCGAACTCGGACCCGACAACACCCTCACCACC 840
              |||||||| |||||| |||||||||| ||||| |||| |||||||  |||||||||  ||
Db        691 CTCCACCACCACGGCATCACCACCTACCTCGAAATCGGCCCCGACACCACCCTCACCGCC 750

Qy        841 CTCACCCACCACAACCTCCCCAACCCCCCCACCACCACCCTCACCCTCACCCACCCCCAC 900
              ||| ||| |  || |||||||| | || ||  |  || || ||||  |  || |  ||||
Db        751 CTCGCCCGCACCACCCTCCCCACCACCACCCACCTCATCCCCCACCACCCGCCGCAACCAC 810

Qy        901 CACCACCCCCAAACC 915
              || |   ||  | |
Db        811 AACGAAGTCCGCAGC 825


RESULT 4
US-09-860-846-1
; Sequence 1, Application US/09860846
; Patent No. US20020164742A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/860,846
;  CURRENT FILING DATE:  2001-05-18
;  PRIOR APPLICATION NUMBER: 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 1
;   LENGTH: 15872
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-860-846-1

  Query Match             25.9%;  Score 239.4;  DB 9;  Length 15872;
  Best Local Similarity   58.0%;  Pred. No. 1.3e-42;
  Matches  523;  Conservative    0;  Mismatches  336;  Indels   42;  Gaps    4;

Qy          1 ACCGCATTCATCTGCTCCGGACAGGGCACCCAACGCCCCGGCATGGCCCACGGCCTCTAC 60
              || || ||| | | | |||| |||||  | || |||  || |||||  | | |  ||   |
Db       4754 ACGGCCTTCCTGTTCACCGGCCAGGGAGCGCAGCGCGTCGCCATGGGCGAGGAACTGCGC 4813
```

```
Qy         61 CACACCCACCCCCGTCTTCGCCGCCGCACTCAACGACATCTGCACCCACCTCGACCCCCAC 120
              | | |||||||| |||||||||||| ||| ||  | | | | |   |||||||| |||
Db       4814 GCCGCGCACCCCGTGTTCGCCGCCGCCCTCGACACCGTGTACGCGGCCCTCGACCGTCAC 4873

Qy        121 CTCGACCACCCCCTCCTCCCCCTCCTCACCCAAAACGACAACGACAACGAGGACGCGGCC 180
              ||||||| , || || |                                || | ||  | ||| |
Db       4874 CTCGACCGGCCGCTGCGG----------------------GAGATCGTCGCCGCCGGG 4909

Qy        181 GCACTGCTCCAGCAGACCCGCTACGCCCAGCCCGCCCTCTTCGCCTTCCAGGTCGCCCTC 240
              |    |||  | |  |||   ||| ||||||||||||||||||||||| |||| || ||
Db       4910 GAGGAGCTGGACCTCACCGCGTACACCCAGCCCGCCCTCTTCGCCTTCGAGGTGGCGCTG 4969

Qy        241 CACCGCCTCCTCACCGACGGCTACCACATCACCCCCCACTACTACGCCGGACACTCCCTC 300
              |||||||||   |||   | ||  | ||  |||| ||      | |||| ||||||  ||
Db       4970 TTCCGCCTCCT---CGAACACCACGGCCTCGTCCCCGACCTGCTCACCGGCCACTCCGTC 5026

Qy        301 GGCGAAATCACCGCCGCCCACCTCGCCGGCATCCTCACCCTCACCGACGCCACCACCCTC 360
              ||||| ||| ||||||| ||| ||||||| |||||  |||||  |||||||  |    |||
Db       5027 GGCGAGATCGCCGCCGCGCACGTCGCCGGTGTCCTCTCCCTCGACGACGCCGCACGTCTC 5086

Qy        361 ATCACCCAACGCGCCACCCTCATGCAAACCATGCCC---CCCGGCACCATGACCACCCTC 417
              |||||   |||| |   |||||||| |     | |     |||| | |||| | || |
Db       5087 GTCACCGCCCGCGGCCGGCTCATGCAGTCGGCCCGCGAGGGCGGCGCGATGATCGCCGTG 5146

Qy        418 CACACCACCCCCCACCACATCACCCACCACCTCACCGCCCACGAAAACGACCTCGCCATC 477
              ||  |  |      | |  ||  | |  ||| |  | | ||||   |     ||||| ||
Db       5147 CAGGCGGGCGAGGCCGAGGTCGTCGAGTCCCTGAAGGGCTACGAGGGCAGGGTCGCCGTC 5206

Qy        478 GCCGCCATCAACACCCCCACCTCCCTCGTCATCAGCGGCACCCCCCACACCGTCCAACAC 537
              |||||| |||||   |||||| || | ||| ||  ||||  | |  || ||| | |  |
Db       5207 GCCGCCGTCAACGGACCCACCGCCGTGGTCGTCTCCGGCGACGCGGACGCCGCCGAGGAG 5266

Qy        538 ATCACCACCCTCTGCCAACAACAAGGCATCAAAACCAAAACCCTCCCCACCAACCACGCC 597
              |||  | || | ||  |||  |||  |  | || | |  || |  | ||  ||| |||||||
Db       5267 ATCCGCGCCGTATGGGCGGGACGCGGCCGGCGCACCCGCAGGCTGCGCGTCAGCCACGCC 5326

Qy        598 TTCCACTCCCCCCACACCAACCCCATCCTCAACCAACTCCACCAGCACACCCAAACCCTC 657
              |||||||||||| ||||   ||  | ||||| || |  ||| || |  | || |   |||
Db       5327 TTCCACTCCCCGCACATGGACGACGTCCTCGACGAGTTCCTCCGGGTCGCCGAGGGCCTG 5386

Qy        658 ACCTACCACCCACCCCACACCCCCCTCATCACCGCCAACAC------------CCCACCC 705
              |||| | |    || |  | |||| || || || |   |||             | |  ||
Db       5387 ACCTTCGAGGAGCCGCGGATCCCCGTCGTCTCCACGGTCACCGGCGCGCTCGTCACGTCC 5446

Qy        706 GACCAACTCCTCACCCCCCACTACTGGACCCAACAAGCCCGCAACACCGTCGACTACGCC 765
              | | | |||  | | |||   ||||||  | | ||  |||    ||||   || |
Db       5447 GGCGAGCTCACCTCGCCCGCGTACTGGGTCGACCAGATCCGGCGGCCCGTGCGCTTCCTG 5506

Qy        766 ACCACCACCCAAACCCTCCACCAACACGGCGTCACCACCTACATCGAACTCGGACCCGAC 825
              | ||  ||  |||||   |   || | ||  |||  |  | ||||  |||| ||||||
Db       5507 GACGCCGTCCGCACCCTGGCCGCCCAGGACGCGACCGTCCTCGTCGAGATCGGCCCCGAC 5566

Qy        826 AACACCCTCACCACCCTCACCCACCACAACCTCCCCAACCCCCCCACCACCACCCTCACC 885
              |  ||||||  | ||| || |  |    ||| |   | | |  | || |   | ||
Db       5567 GCCGTCCTCACGGCACTCGCCGAGGAGGCTCTCGCGCCCGGCACGGACGCCCCGGACGCC 5626

Qy        886 C 886
              |
Db       5627 C 5627


RESULT 5
US-09-988-384B-1
; Sequence 1, Application US/09988384B
; Publication No. US20030073824A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.536US1
;  CURRENT APPLICATION NUMBER: US/09/988,384B
;  CURRENT FILING DATE:  2001-11-19
;  PRIOR APPLICATION NUMBER: PCT/US99/14398
;  PRIOR FILING DATE: 1999-06-25
;  PRIOR APPLICATION NUMBER: US-09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 53
; SEQ ID NO 1
;   LENGTH: 15872
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-988-384B-1

  Query Match             25.9%;  Score 239.4;  DB 9;  Length 15872;
  Best Local Similarity   58.0%;  Pred. No. 1.3e-42;
  Matches  523;  Conservative    0;  Mismatches  336;  Indels   42;  Gaps    4;

Qy          1 ACCGCATTCATCTGCTCCGGACAGGGCACCCAACGCCCCGGCATGGCCCACGGCCTCTAC 60
              || || ||| | | | |||| |||||  | || |||  || |||||  | | |  ||   |
Db       4754 ACGGCCTTCCTGTTCACCGGCCAGGGAGCGCAGCGCGTCGCCATGGGCGAGGAACTGCGC 4813

Qy         61 CACACCCACCCCGTCTTCGCCGCCGCACTCAACGACATCTGCACCCACCTCGACCCCCAC 120
              | | |||||||| |||||||||| ||| ||  | | | | |   |||||||| |||
Db       4814 GCCGCGCACCCCGTGTTCGCCGCCGCCCTCGACACCGTGTACGCGGCCCTCGACCGTCAC 4873

Qy        121 CTCGACCACCCCCTCCTCCCCCTCCTCACCCAAAACGACAACGACAACGAGGACGCGGCC 180
              |||||||  || || |                                || | ||  | ||| |
Db       4874 CTCGACCGGCCGCTGCGG----------------------GAGATCGTCGCCGCCGGG 4909

Qy        181 GCACTGCTCCAGCAGACCCGCTACGCCCAGCCCGCCCTCTTCGCCTTCCAGGTCGCCCTC 240
              |    |||  | |  |||   ||| ||||||||||||||||||||||| |||| || ||
Db       4910 GAGGAGCTGGACCTCACCGCGTACACCCAGCCCGCCCTCTTCGCCTTCGAGGTGGCGCTG 4969

Qy        241 CACCGCCTCCTCACCGACGGCTACCACATCACCCCCCACTACTACGCCGGACACTCCCTC 300
              |||||||||   |||   | ||  | ||  |||| ||      | |||| ||||||  ||
Db       4970 TTCCGCCTCCT---CGAACACCACGGCCTCGTCCCCGACCTGCTCACCGGCCACTCCGTC 5026

Qy        301 GGCGAAATCACCGCCGCCCACCTCGCCGGCATCCTCACCCTCACCGACGCCACCACCCTC 360
              ||||| ||| ||||||| ||| ||||||| |||||  |||||  |||||||  |    |||
Db       5027 GGCGAGATCGCCGCCGCGCACGTCGCCGGTGTCCTCTCCCTCGACGACGCCGCACGTCTC 5086

Qy        361 ATCACCCAACGCGCCACCCTCATGCAAACCATGCCC---CCCGGCACCATGACCACCCTC 417
              |||||   |||| |   |||||||| |     | |     |||| | |||| | || |
Db       5087 GTCACCGCCCGCGGCCGGCTCATGCAGTCGGCCCGCGAGGGCGGCGCGATGATCGCCGTG 5146

Qy        418 CACACCACCCCCCACCACATCACCCACCACCTCACCGCCCACGAAAACGACCTCGCCATC 477
              ||  |  |      | |  ||  | |  ||| |  | | ||||   |     ||||| ||
Db       5147 CAGGCGGGCGAGGCCGAGGTCGTCGAGTCCCTGAAGGGCTACGAGGGCAGGGTCGCCGTC 5206

Qy        478 GCCGCCATCAACACCCCCACCTCCCTCGTCATCAGCGGCACCCCCCACACCGTCCAACAC 537
              |||||| |||||   |||||| || | ||| ||  ||||  | |  || ||| | |  |
Db       5207 GCCGCCGTCAACGGACCCACCGCCGTGGTCGTCTCCGGCGACGCGGACGCCGCCGAGGAG 5266

Qy        538 ATCACCACCCTCTGCCAACAACAAGGCATCAAAACCAAAACCCTCCCCACCAACCACGCC 597
              |||  | || | ||  |||  |||  |  | || | |  || |  | ||  ||| |||||||
Db       5267 ATCCGCGCCGTATGGGCGGGACGCGGCCGGCGCACCCGCAGGCTGCGCGTCAGCCACGCC 5326

Qy        598 TTCCACTCCCCCCACACCAACCCCATCCTCAACCAACTCCACCAGCACACCCAAACCCTC 657
              |||||||||||| ||||   ||  | ||||| || |  ||| || |  | || |   |||
Db       5327 TTCCACTCCCCGCACATGGACGACGTCCTCGACGAGTTCCTCCGGGTCGCCGAGGGCCTG 5386

Qy        658 ACCTACCACCCACCCCACACCCCCCTCATCACCGCCAACAC------------CCCACCC 705
              |||| | |    || |  | |||| || || || |   |||             | |  ||
Db       5387 ACCTTCGAGGAGCCGCGGATCCCCGTCGTCTCCACGGTCACCGGCGCGCTCGTCACGTCC 5446

Qy        706 GACCAACTCCTCACCCCCCACTACTGGACCCAACAAGCCCGCAACACCGTCGACTACGCC 765
              | | | |||  | | |||   ||||||  | | ||  |||    ||||   || |
Db       5447 GGCGAGCTCACCTCGCCCGCGTACTGGGTCGACCAGATCCGGCGGCCCGTGCGCTTCCTG 5506

Qy        766 ACCACCACCCAAACCCTCCACCAACACGGCGTCACCACCTACATCGAACTCGGACCCGAC 825
              | ||  ||  |||||   |   || | ||  |||  |  | ||||  |||| ||||||
Db       5507 GACGCCGTCCGCACCCTGGCCGCCCAGGACGCGACCGTCCTCGTCGAGATCGGCCCCGAC 5566
```

```
Qy        826 AACACCCTCACCACCCTCACCCACCACAACCTCCCCAACCCCCCCACCACCACCCTCACC 885
              |  ||||||  | ||| || |  |   ||| |   |  | |   | || |   | ||
Db       5567 GCCGTCCTCACGGCACTCGCCGAGGAGGCTCTCGCGCCCGGCACGGACGCCCCGGACGCC 5626

Qy        886 C 886
              |
Db       5627 C 5627


RESULT 6
US-09-836-821-1
; Sequence 1, Application US/09836821
; Publication No. US20030087405A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/836,821
;  CURRENT FILING DATE:  2001-04-17
;  PRIOR APPLICATION NUMBER: 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 1
;   LENGTH: 15872
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-836-821-1

  Query Match             25.9%;  Score 239.4;  DB 9;  Length 15872;
  Best Local Similarity   58.0%;  Pred. No. 1.3e-42;
  Matches  523;  Conservative    0;  Mismatches  336;  Indels   42;  Gaps    4;

Qy          1 ACCGCATTCATCTGCTCCGGACAGGGCACCCAACGCCCCGGCATGGCCCACGGCCTCTAC 60
              || || ||| | | | |||| |||||  | || |||  || ||||| | | |  ||   |
Db       4754 ACGGCCTTCCTGTTCACCGGCCAGGGAGCGCAGCGCGTCGCCATGGGCGAGGAACTGCGC 4813

Qy         61 CACACCCACCCCGTCTTCGCCGCCGCACTCAACGACATCTGCACCCACCTCGACCCCCAC 120
                | | |||||||| |||||||||||| ||| ||  | | | |   |||||||  |||
Db       4814 GCCGCGCACCCCGTGTTCGCCGCCGCCCTCGACACCGTGTACGCGGCCCTCGACCGTCAC 4873

Qy        121 CTCGACCACCCCCTCCTCCCCCTCCTCACCCAAAACGACAACGACAACGAGGACGCGGCC 180
              |||||||  || || |                          || | ||  | ||| |
Db       4874 CTCGACCGGCCGCTGCGG----------------------GAGATCGTCGCCGCCGGG 4909

Qy        181 GCACTGCTCCAGCAGACCCGCTACGCCCAGCCCGCCCTCTTCGCCTTCCAGGTCGCCCTC 240
              |   |||  | |  |||  ||| ||||||||||||||||||||||| |||| || ||
Db       4910 GAGGAGCTGGACCTCACCGCGTACACCCAGCCCGCCCTCTTCGCCTTCGAGGTGGCGCTG 4969

Qy        241 CACCGCCTCCTCACCGACGGCTACCACATCACCCCCCACTACTACGCCGGACACTCCCTC 300
              |||||||||   |||   | ||  | |  |||| ||      | |||| ||||| ||
Db       4970 TTCCGCCTCCT---CGAACACCACGGCCTCGTCCCCGACCTGCTCACCGGCCACTCCGTC 5026

Qy        301 GGCGAAATCACCGCCGCCCACCTCGCCGGCATCCTCACCCTCACCGACGCCACCACCCTC 360
              ||||| ||| ||||||| ||| ||||||| |||||  |||||  ||||||| |    |||
Db       5027 GGCGAGATCGCCGCCGCGCACGTCGCCGGTGTCCTCTCCCTCGACGACGCCGCACGTCTC 5086

Qy        361 ATCACCCAACGCGCCACCCTCATGCAAACCATGCCC---CCCGGCACCATGACCACCCTC 417
              |||||   |||| |   |||||||   |    | |    |||| | |||| | || |
Db       5087 GTCACCGCCCGCGGCCGGCTCATGCAGTCGGCCCGCGAGGGCGGCGCGATGATCGCCGTG 5146

Qy        418 CACACCACCCCCCACCACATCACCCACCACCTCACCGCCCACGAAAACGACCTCGCCATC 477
              ||  |   |      | |  ||  | |   ||| |  | | ||||    |    ||||| ||
Db       5147 CAGGCGGGCGAGGCCGAGGTCGTCGAGTCCCTGAAGGGCTACGAGGGCAGGGTCGCCGTC 5206

Qy        478 GCCGCCATCAACACCCCCACCTCCCTCGTCATCAGCGGCACCCCCCACACCGTCCAACAC 537
              |||||| |||||   |||||| || | ||| ||  ||||  | |  || ||| | |  |
Db       5207 GCCGCCGTCAACGGACCCACCGCCGTGGTCGTCTCCGGCGACGCGGACGCCGCCGAGGAG 5266

Qy        538 ATCACCACCCTCTGCCAACAACAAGGCATCAAAACCAAAACCCTCCCCACCAACCACGCC 597
                |||  | || | ||       ||  |||      |||   |  || | |  || |||||||
Db       5267 ATCCGCGCCGTATGGGCGGGACGCGGCCGGCGCACCCGCAGGCTGCGCGTCAGCCACGCC 5326

Qy        598 TTCCACTCCCCCCACACCAACCCCATCCTCAACCAACTCCACCAGCACACCCAAACCCTC 657
              |||||||||||| ||||   ||  | ||||| || |  ||| || |  | || |   |||
Db       5327 TTCCACTCCCCGCACATGGACGACGTCCTCGACGAGTTCCTCCGGGTCGCCGAGGGCCTG 5386

Qy        658 ACCTACCACCCACCCCACACCCCCCTCATCACCGCCAACAC------------CCCACCC 705
              |||| | |    || |  | |||| || || || |   |||             | |  ||
Db       5387 ACCTTCGAGGAGCCGCGGATCCCCGTCGTCTCCACGGTCACCGGCGCGCTCGTCACGTCC 5446

Qy        706 GACCAACTCCTCACCCCCCACTACTGGACCCAACAAGCCCGCAACACCGTCGACTACGCC 765
              | | | |||  | | |||   ||||||  | | ||   |||      ||||   || | |
Db       5447 GGCGAGCTCACCTCGCCCGCGTACTGGGTCGACCAGATCCGGCGGCCCCGTGCGCTTCCTG 5506

Qy        766 ACCACCACCCAAACCCTCCACCAACACGGCGTCACCACCTACATCGAACTCGGACCCGAC 825
              | ||  ||  |||||   |   || | ||  |||  |  | ||||  |||| ||||||
Db       5507 GACGCCGTCCGCACCCTGGCCGCCCAGGACGCGACCGTCCTCGTCGAGATCGGCCCCGAC 5566

Qy        826 AACACCCTCACCACCCTCACCCACCACAACCTCCCCAACCCCCCCACCACCACCCTCACC 885
              |  ||||||  | ||| || |  |   ||| |   |  | |   | || |   | ||
Db       5567 GCCGTCCTCACGGCACTCGCCGAGGAGGCTCTCGCGCCCGGCACGGACGCCCCGGACGCC 5626

Qy        886 C 886
              |
Db       5627 C 5627


RESULT 7
US-09-861-289-1
; Sequence 1, Application US/09861289
; Patent No. US20020110897A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/861,289
;  CURRENT FILING DATE:  2001-05-18
;  PRIOR APPLICATION NUMBER: 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 1
;   LENGTH: 15872
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-861-289-1

  Query Match             25.9%;  Score 239.4;  DB 10;  Length 15872;
  Best Local Similarity   58.0%;  Pred. No. 1.3e-42;
  Matches  523;  Conservative    0;  Mismatches  336;  Indels   42;  Gaps    4;

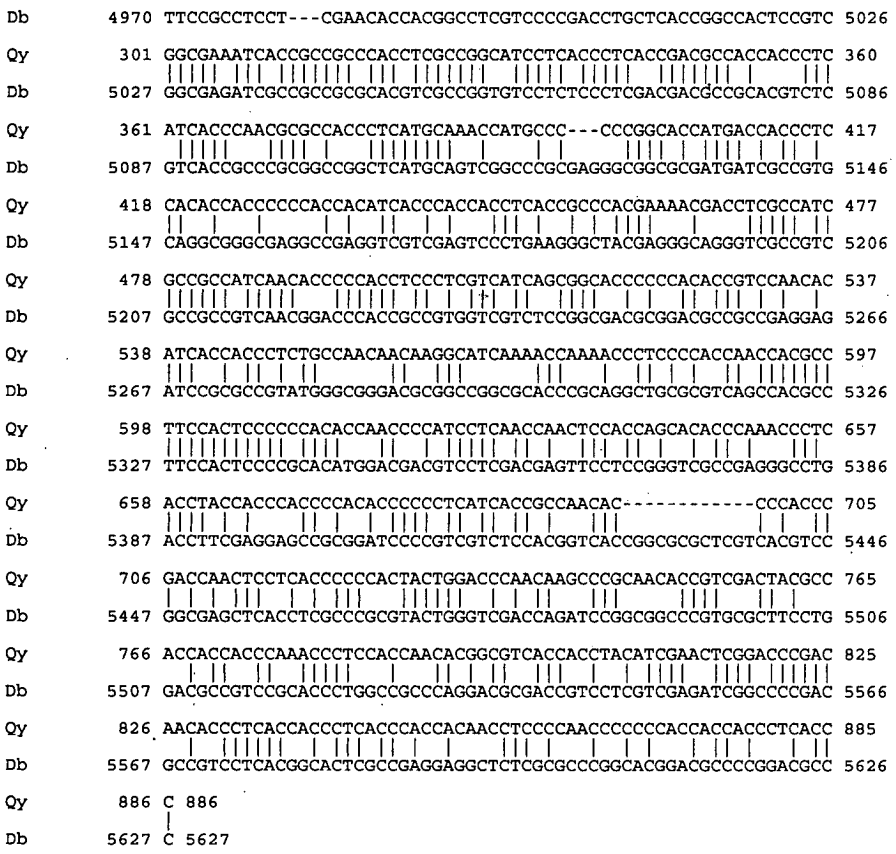
Qy          1 ACCGCATTCATCTGCTCCGGACAGGGCACCCAACGCCCCGGCATGGCCCACGGCCTCTAC 60
              || || ||| | | | |||| |||||  | || |||  || ||||| | | |  ||   |
Db       4754 ACGGCCTTCCTGTTCACCGGCCAGGGAGCGCAGCGCGTCGCCATGGGCGAGGAACTGCGC 4813

Qy         61 CACACCCACCCCGTCTTCGCCGCCGCACTCAACGACATCTGCACCCACCTCGACCCCCAC 120
                | | |||||||| |||||||||||| ||| ||  | | | |   |||||||  |||
Db       4814 GCCGCGCACCCCGTGTTCGCCGCCGCCCTCGACACCGTGTACGCGGCCCTCGACCGTCAC 4873

Qy        121 CTCGACCACCCCCTCCTCCCCCTCCTCACCCAAAACGACAACGACAACGAGGACGCGGCC 180
              |||||||  || || |                          || | ||  | ||| |
Db       4874 CTCGACCGGCCGCTGCGG----------------------GAGATCGTCGCCGCCGGG 4909

Qy        181 GCACTGCTCCAGCAGACCCGCTACGCCCAGCCCGCCCTCTTCGCCTTCCAGGTCGCCCTC 240
              |   |||  | |  |||  ||| ||||||||||||||||||||||| |||| || ||
Db       4910 GAGGAGCTGGACCTCACCGCGTACACCCAGCCCGCCCTCTTCGCCTTCGAGGTGGCGCTG 4969

Qy        241 CACCGCCTCCTCACCGACGGCTACCACATCACCCCCCACTACTACGCCGGACACTCCCTC 300
              |||||||||   |||   | ||  | |  |||| ||      | |||| ||||| ||
```



RESULT 8
US-09-808-880-1
; Sequence 1, Application US/09808880
; Publication No. US20030027287A1
; GENERAL INFORMATION:
; APPLICANT: Betlach, Mary C.
; APPLICANT: Shah, Sanjay Krishnakant
; APPLICANT: McDaniel, Robert
; APPLICANT: Tang, Li
; TITLE OF INVENTION: RECOMBINANT OLEANDOLIDE POLYKETIDE SYNTHASE
; FILE REFERENCE: 30062-20029.00
; CURRENT APPLICATION NUMBER: US/09/808,880
; CURRENT FILING DATE: 2001-03-14
; PRIOR APPLICATION NUMBER: US/09/428,517
; PRIOR FILING DATE: 1999-10-28
; PRIOR APPLICATION NUMBER: 60/120,254
; PRIOR FILING DATE: 1999-02-16
; PRIOR APPLICATION NUMBER: 60/106,100
; PRIOR FILING DATE: 1998-10-29
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 50937
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artifical Sequence: Recombinant DNA
US-09-808-880-1

Query Match 25.2%; Score 232.6; DB 9; Length 50937;
Best Local Similarity 57.8%; Pred. No. 3.6e-41;
Matches 528; Conservative 0; Mismatches 349; Indels 36; Gaps 5;

RESULT 9

```
US-09-860-846-30
; Sequence 30, Application US/09860846
; Patent No. US20020164742A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/860,846
;  CURRENT FILING DATE:  2001-05-18
;  PRIOR APPLICATION NUMBER: 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 30
;   LENGTH: 13842
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-860-846-30

  Query Match             23.8%;  Score 220.2;  DB 9;  Length 13842;
  Best Local Similarity   56.5%;  Pred. No. 1.6e-38;
  Matches  527;  Conservative    0;  Mismatches  373;  Indels   33;  Gaps    5;

Qy          1 ACCGCATTCATCTGCTCCGGACAGGGCACCCAACGCCCCGGCATGGCCCACGGCCTCTAC 60
              ||||| ||| ||| | |||| |||||||  || |||||||||   | ||  |  || |||
Db       9532 ACCGCGTTCCTCTTCACCGGCCAGGGCAGTCAGCGCCCCGGCGCCGGCCGTGAGCTGTAC 9591

Qy         61 CACACCCACCCCGTCTTCGCCGCCGCACTCAACGACATCTGCACCCACCTCGACCCCCAC 120
              ||   ||||||||||||||||   || ||| |||| |||||| ||||||||||   |||
Db       9592 GACCGGCACCCCGTCTTCGCCCGGGCGCTCGACGAGATCTGCGCCCACCTCGACGGTCAC 9651

Qy        121 CTCGACCACCCCCTCCTCCCCCTCCTCACCCAAAACGACAACGACAACGAGGACGCGGCC 180
              ||||| |  ||||| |||  | |  |          ||     |||    |   | ||||
Db       9652 CTCGAACTGCCCCTGCTCGACGTGAT------GTTCGCGGCCGAGGGCAGCGCGGAGGCC 9705

Qy        181 GCACTGCTCCAGCAGACCCGCTACGCCCAGCCCGCCCTCTTCGCCTTCCAGGTCGCCCTC 240
              || |||||| |  |||| || ||| | |||  ||| || ||||||  |  |||||| |||
Db       9706 GCGCTGCTCGACGAGACGCGGTACACGCAGTGCGCGCTGTTCGCCCTGGAGGTCGCGCTC 9765

Qy        241 CACCGCCTCCTCACCGACGGCTACCACATCACCCCCCACTACTACGCCGGACACTCCCTC 300
              |||  ||| |   |||  |||    |||     ||   |         ||| |||||  ||
Db       9766 TTCCGGCTCGT---CGAGAGCTGGGGCATGCGGCCGGCCGCACTGCTCGGTCACTCGGTC 9822

Qy        301 GGCGAAATCACCGCCGCCCACCTCGCCGGCATCCTCACCCTCACCGACGCCACCACCCTC 360
              ||||| ||| ||||||| ||| |||||||   | || | ||| |||||||| ||  |||
Db       9823 GGCGAGATCGCCGCCGCGCACGTCGCCGGTGTGTTCTCGCTCGCCGACGCCGCCCGCCTG 9882

Qy        361 ATCACCCAACGCGCCACCCTCATGCAAACCATGCCCCCC---GGCACCATGACCACCCTC 417
              || ||   |||| |   ||||||||     ||||| ||   ||| | ||| | || ||
Db       9883 GTCGCCGCGCGCGGCCGGCTCATGCAGGAGCTGCCCGCCGGTGGCGCGATGCTCGCCGTC 9942

Qy        418 CACACCACCCCCCACCACATCACCCACCACCTCACCGCCCACGAAAACGAC--------- 468
              ||  ||  |     || | |||  |     ||     |  | ||     ||
Db       9943 CAGGCCGCGGAGGACGAGATCCGCGTGTGGCTGGAGACGGAGGAGCGGTACGCGGGACGT 10002

Qy        469 CTCGCCATCGCCGCCATCAACACCCCCACCTCCCTCGTCATCAGCGGCACCCCCCACACC 528
              || | | |||||||| |||||  ||||     ||  |||| |   ||||  | |  || |
Db       10003 CTGGACGTCGCCGCCGTCAACGGCCCCGAGGCCGCCGTCCTGTCCGGCGACGCGGACGCG 10062

Qy        529 GTCCAACACATCACCACCCTCTGCCAACAACAAGGCATCAAAACCAAAACCCTCCCCACC 588
              | |  |          |  |||     | |||  ||  |||     | || |    |
Db       10063 GCGCGGGAGGCGGAGGCGTACTGGTCCGGGCTCGGCCGCAGGACCCGCGCGCTGCGGGTC 10122

Qy        589 AACCACGCCTTCCACTCCCCCCACACCAACCCCATCCTCAACCAACTCCACCAGCACACC 648
              | |||||||||||||||| | ||||   ||  ||| ||| ||     ||| |     |
Db       10123 AGCCACGCCTTCCACTCCGCGCACATGGACGGCATGCTCGACGGGTTCCGCGCCGTCCTG 10182

Qy        649 CAAACCCTCACCTACCACCCACCCCACACCCCCCTCATCACCGCCAACA---------- 697
              | |||  |   | ||  |  |||  |    || |  || |   |  ||
Db       10183 GAGACGGTGGAGTTCCGGCGCCCCTCCCTGACCGTGGTCTCGAACGTCACCGGCCTGGCC 10242

Qy        698 -CCCCACCCGACCAACTCCTCACCCCCCACTACTGGACCCAACAAGCCCGCAACACCGTC 756
               ||   || ||| | ||    | |||| | |||||||  ||  || | ||||  |||||||
Db       10243 GCCGGCCCGGACGACCTGTGCGACCCCGAGTACTGGGTCCGGCACGTCCGCGGCACCGTC 10302

Qy        757 GACTACGCCACCACCACCCAAACCCTCCACCAACACGGCGTCACCACCTACATCGAACTC 816
              || |  |  |  |  ||    ||| | | | | ||||||    |||| | | || ||
Db       10303 CGCTTCCTCGACGGCGTCCGTGTCCTGCGCGACCTCGGCGTGCGGACCTGCCTGGAGCTG 10362

Qy        817 GGACCCGACAACACCCTCACCACCCTCACCCACCACAACCTCCCCAACCCCCCCACCACC 876
              || ||||||     |||||||  ||  |   |  ||  |||| |  ||  ||||| |
Db       10363 GGCCCCGACGGGGTCCTCACCGCCATGGCGGCCGACGGCCTCGCGGACACCCCCGCGGAT 10422

Qy        877 ACCCTCACCCTCACCCACCCCCACCACCACCCC 909
              ||    ||  | ||| |  |  |   | | ||
Db       10423 TCCGCTGCCGGCTCCCCCGTCGGCTCTCCCGCC 10455


RESULT 10
US-09-988-384B-30
; Sequence 30, Application US/09988384B
; Publication No. US20030073824A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.536US1
;  CURRENT APPLICATION NUMBER: US/09/988,384B
;  CURRENT FILING DATE:  2001-11-19
;  PRIOR APPLICATION NUMBER: PCT/US99/14398
;  PRIOR FILING DATE: 1999-06-25
;  PRIOR APPLICATION NUMBER: US 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 53
; SEQ ID NO 30
;   LENGTH: 13842
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-988-384B-30

  Query Match             23.8%;  Score 220.2;  DB 9;  Length 13842;
  Best Local Similarity   56.5%;  Pred. No. 1.6e-38;
  Matches  527;  Conservative    0;  Mismatches  373;  Indels   33;  Gaps    5;

Qy          1 ACCGCATTCATCTGCTCCGGACAGGGCACCCAACGCCCCGGCATGGCCCACGGCCTCTAC 60
              ||||| ||| ||| | |||| |||||||  || |||||||||   | ||  |  || |||
Db       9532 ACCGCGTTCCTCTTCACCGGCCAGGGCAGTCAGCGCCCCGGCGCCGGCCGTGAGCTGTAC 9591

Qy         61 CACACCCACCCCGTCTTCGCCGCCGCACTCAACGACATCTGCACCCACCTCGACCCCCAC 120
              ||   ||||||||||||||||   || ||| |||| |||||| ||||||||||   |||
Db       9592 GACCGGCACCCCGTCTTCGCCCGGGCGCTCGACGAGATCTGCGCCCACCTCGACGGTCAC 9651

Qy        121 CTCGACCACCCCCTCCTCCCCCTCCTCACCCAAAACGACAACGACAACGAGGACGCGGCC 180
              ||||| |  ||||| |||  | |  |          ||     |||    |   | ||||
Db       9652 CTCGAACTGCCCCTGCTCGACGTGAT------GTTCGCGGCCGAGGGCAGCGCGGAGGCC 9705

Qy        181 GCACTGCTCCAGCAGACCCGCTACGCCCAGCCCGCCCTCTTCGCCTTCCAGGTCGCCCTC 240
              || |||||| |  |||| || ||| | |||  ||| || ||||||  |  |||||| |||
Db       9706 GCGCTGCTCGACGAGACGCGGTACACGCAGTGCGCGCTGTTCGCCCTGGAGGTCGCGCTC 9765

Qy        241 CACCGCCTCCTCACCGACGGCTACCACATCACCCCCCACTACTACGCCGGACACTCCCTC 300
              |||  ||| |   |||  |||    |||     ||   |         ||| |||||  ||
Db       9766 TTCCGGCTCGT---CGAGAGCTGGGGCATGCGGCCGGCCGCACTGCTCGGTCACTCGGTC 9822

Qy        301 GGCGAAATCACCGCCGCCCACCTCGCCGGCATCCTCACCCTCACCGACGCCACCACCCTC 360
              ||||| ||| ||||||| ||| |||||||   | || | ||| |||||||| ||  |||
Db       9823 GGCGAGATCGCCGCCGCGCACGTCGCCGGTGTGTTCTCGCTCGCCGACGCCGCCCGCCTG 9882

Qy        361 ATCACCCAACGCGCCACCCTCATGCAAACCATGCCCCCC---GGCACCATGACCACCCTC 417
              || ||   |||| |   ||||||||     ||||| ||   ||| | ||| | || ||
Db       9883 GTCGCCGCGCGCGGCCGGCTCATGCAGGAGCTGCCCGCCGGTGGCGCGATGCTCGCCGTC 9942
```

```
Qy        418 CACACCACCCCCCACCACATCACCCACCACCTCACCGCCCACGAAAACGAC--------- 468
              ||  || |     || | |||  |     ||     |  | ||     ||
Db       9943 CAGGCCGCGGAGGACGAGATCCGCGTGTGGCTGGAGACGGAGGAGCGGTACGCGGGACGT 10002

Qy        469 CTCGCCATCGCCGCCATCAACACCCCCACCTCCCTCGTCATCAGCGGCACCCCCCACACC 528
              || | | |||||||| |||||  ||||    ||  |||| |   ||||  | |  || |
Db      10003 CTGGACGTCGCCGCCGTCAACGGCCCCGAGGCCGCCGTCCTGTCCGGCGACGCGGACGCG 10062

Qy        529 GTCCAACACATCACCACCCTCTGCCAACAACAAGGCATCAAAACCAAAACCCTCCCCACC 588
              | |   |          |   |||         | |||  ||  |||     | || |    |
Db      10063 GCGCGGGAGGCGGAGGCGTACTGGTCCGGGCTCGGCCGCAGGACCCGCGCGCTGCGGGTC 10122

Qy        589 AACCACGCCTTCCACTCCCCCCACACCAACCCCATCCTCAACCAACTCCACCAGCACACC 648
              | |||||||||||||||| | ||||   ||  ||| ||| ||    ||| |      |
Db      10123 AGCCACGCCTTCCACTCCGCGCACATGGACGGCATGCTCGACGGGTTCCGCGCCGTCCTG 10182

Qy        649 CAAACCCTCACCTACCACCCACCCCACACCCCCCTCATCACCGCCAACA----------- 697
              | ||  |    | ||  |  |||  |     || |  || |   |  ||
Db      10183 GAGACGGTGGAGTTCCGGCGCCCCTCCCTGACCGTGGTCTCGAACGTCACCGGCCTGGCC 10242

Qy        698 -CCCCACCCGACCAACTCCTCACCCCCCACTACTGGACCCAACAAGCCCGCAACACCGTC 756
               ||   || ||| | ||    | |||| | ||||||  ||  || | ||||  |||||||
Db      10243 GCCGGCCCGGACGACCTGTGCGACCCCGAGTACTGGGTCCGGCACGTCCGCGGCACCGTC 10302

Qy        757 GACTACGCCACCACCACCCAAACCCTCCACCAACACGGCGTCACCACCTACATCGAACTC 816
              || |  |  |  |  ||    ||| | | | | ||||||    |||| | |  || ||
Db      10303 CGCTTCCTCGACGGCGTCCGTGTCCTGCGCGACCTCGGCGTGCGGACCTGCCTGGAGCTG 10362

Qy        817 GGACCCGACAACACCCTCACCACCCTCACCCACCACAACCTCCCCAACCCCCCCACCACC 876
              || ||||||     ||||||| || |  |    | ||  |||| |  || ||||| |
Db      10363 GGCCCCGACGGGGTCCTCACCGCCATGGCGGCCGACGGCCTCGCGGACACCCCCGCGGAT 10422

Qy        877 ACCCTCACCCTCACCCACCCCCACCACCACCCC 909
              ||    ||  | ||| |  |  |   | | ||
Db      10423 TCCGCTGCCGGCTCCCCCGTCGGCTCTCCCGCC 10455


RESULT 11
US-09-836-821-30
; Sequence 30, Application US/09836821
; Publication No. US20030087405A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/836,821
;  CURRENT FILING DATE:  2001-04-17
;  PRIOR APPLICATION NUMBER: 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 30
;   LENGTH: 13842
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-836-821-30

  Query Match             23.8%;  Score 220.2;  DB 9;  Length 13842;
  Best Local Similarity   56.5%;  Pred. No. 1.6e-38;
  Matches  527;  Conservative    0;  Mismatches  373;  Indels   33;  Gaps    5;

Qy          1 ACCGCATTCATCTGCTCCGGACAGGGCACCCAACGCCCCGGCATGGCCCACGGCCTCTAC 60
              ||||| ||| ||| | |||| |||||||  || |||||||||    | ||  |  || |||
Db       9532 ACCGCGTTCCTCTTCACCGGCCAGGGCAGTCAGCGCCCCGGCGCCGGCCGTGAGCTGTAC 9591

Qy         61 CACACCCACCCCGTCTTCGCCGCCGCACTCAACGACATCTGCACCCACCTCGACCCCCAC 120
               ||   |||||||||||||||   || ||| |||| |||||| ||||||||||    |||
Db       9592 GACCGGCACCCCGTCTTCGCCCGGGCGCTCGACGAGATCTGCGCCCACCTCGACGGTCAC 9651

Qy        121 CTCGACCACCCCCTCCTCCCCCTCCTCACCCAAAACGACAACGACAACGAGGACGCGGCC 180
              ||||| |  ||||| |||  | |  |          ||    |||  |   |  | ||||
Db       9652 CTCGAACTGCCCCTGCTCGACGTGAT------GTTCGCGGCCGAGGGCAGCGCGGAGGCC 9705

Qy        181 GCACTGCTCCAGCAGACCCGCTACGCCCAGCCCGCCCTCTTCGCCTTCCAGGTCGCCCTC 240
              || |||||| |  |||| || ||| | |||  ||| || |||||| |  ||||||| |||
Db       9706 GCGCTGCTCGACGAGACGCGGTACACGCAGTGCGCGCTGTTCGCCCTGGAGGTCGCGCTC 9765

Qy        241 CACCGCCTCCTCACCGACGGCTACCACATCACCCCCCACTACTACGCCGGACACTCCCTC 300
              |||  ||| |   |||  |||    |||    ||  |          ||| |||||  ||
Db       9766 TTCCGGCTCGT---CGAGAGCTGGGGCATGCGGCCGGCCGCACTGCTCGGTCACTCGGTC 9822

Qy        301 GGCGAAATCACCGCCGCCCACCTCGCCGGCATCCTCACCCTCACCGACGCCACCACCCTC 360
              ||||| ||| ||||||| ||| |||||||  |  || | ||| |||||||||  ||  |||
Db       9823 GGCGAGATCGCCGCCGCGCACGTCGCCGGTGTGTTCTCGCTCGCCGACGCCGCCCGCCTG 9882

Qy        361 ATCACCCAACGCGCCACCCTCATGCAAACCATGCCCCCC---GGCACCATGACCACCCTC 417
              || ||   |||| |  ||||||||      ||||| ||   ||| | |||  | || ||
Db       9883 GTCGCCGCGCGCGGCCGGCTCATGCAGGAGCTGCCCGCCGGTGGCGCGATGCTCGCCGTC 9942

Qy        418 CACACCACCCCCCACCACATCACCCACCACCTCACCGCCCACGAAAACGAC--------- 468
              ||  || |     || | |||  |     ||     |  | ||     ||
Db       9943 CAGGCCGCGGAGGACGAGATCCGCGTGTGGCTGGAGACGGAGGAGCGGTACGCGGGACGT 10002

Qy        469 CTCGCCATCGCCGCCATCAACACCCCCACCTCCCTCGTCATCAGCGGCACCCCCCACACC 528
              || | | |||||||| |||||  ||||    ||  |||| |   ||||  | |  || |
Db      10003 CTGGACGTCGCCGCCGTCAACGGCCCCGAGGCCGCCGTCCTGTCCGGCGACGCGGACGCG 10062

Qy        529 GTCCAACACATCACCACCCTCTGCCAACAACAAGGCATCAAAACCAAAACCCTCCCCACC 588
              | |   |          |   |||         | |||  ||  |||     | || |    |
Db      10063 GCGCGGGAGGCGGAGGCGTACTGGTCCGGGCTCGGCCGCAGGACCCGCGCGCTGCGGGTC 10122

Qy        589 AACCACGCCTTCCACTCCCCCCACACCAACCCCATCCTCAACCAACTCCACCAGCACACC 648
              | |||||||||||||||| | ||||   ||  ||| ||| ||    ||| |      |
Db      10123 AGCCACGCCTTCCACTCCGCGCACATGGACGGCATGCTCGACGGGTTCCGCGCCGTCCTG 10182

Qy        649 CAAACCCTCACCTACCACCCACCCCACACCCCCCTCATCACCGCCAACA----------- 697
              | ||  |    | ||  |  |||  |     || |  || |   |  ||
Db      10183 GAGACGGTGGAGTTCCGGCGCCCCTCCCTGACCGTGGTCTCGAACGTCACCGGCCTGGCC 10242

Qy        698 -CCCCACCCGACCAACTCCTCACCCCCCACTACTGGACCCAACAAGCCCGCAACACCGTC 756
               ||   || ||| | ||    | |||| | ||||||  ||  || | ||||  |||||||
Db      10243 GCCGGCCCGGACGACCTGTGCGACCCCGAGTACTGGGTCCGGCACGTCCGCGGCACCGTC 10302

Qy        757 GACTACGCCACCACCACCCAAACCCTCCACCAACACGGCGTCACCACCTACATCGAACTC 816
              || |  |  |  |  ||    ||| | | | | ||||||    |||| | |  || ||
Db      10303 CGCTTCCTCGACGGCGTCCGTGTCCTGCGCGACCTCGGCGTGCGGACCTGCCTGGAGCTG 10362

Qy        817 GGACCCGACAACACCCTCACCACCCTCACCCACCACAACCTCCCCAACCCCCCCACCACC 876
              || ||||||     ||||||| || |  |    | ||  |||| |  || ||||| |
Db      10363 GGCCCCGACGGGGTCCTCACCGCCATGGCGGCCGACGGCCTCGCGGACACCCCCGCGGAT 10422

Qy        877 ACCCTCACCCTCACCCACCCCCACCACCACCCC 909
              ||    ||  | ||| |  |  |   | | ||
Db      10423 TCCGCTGCCGGCTCCCCCGTCGGCTCTCCCGCC 10455


RESULT 12
US-09-861-289-30
; Sequence 30, Application US/09861289
; Patent No. US20020110897A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/861,289
;  CURRENT FILING DATE:  2001-05-18
;  PRIOR APPLICATION NUMBER: 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
```

```
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 30
;   LENGTH: 13842
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-861-289-30

  Query Match             23.8%;  Score 220.2;  DB 10;  Length 13842;
  Best Local Similarity   56.5%;  Pred. No. 1.6e-38;
  Matches  527;  Conservative    0;  Mismatches  373;  Indels   33;  Gaps    5;

Qy          1 ACCGCATTCATCTGCTCCGGACAGGGCACCCAACGCCCCGGCATGGCCCACGGCCTCTAC 60
              ||||| ||| ||| | |||| |||||||  || |||||||||   | ||  |  || |||
Db       9532 ACCGCGTTCCTCTTCACCGGCCAGGGCAGTCAGCGCCCCGGCGCCGGCCGTGAGCTGTAC 9591

Qy         61 CACACCCACCCCGTCTTCGCCGCCGCACTCAACGACATCTGCACCCACCTCGACCCCCAC 120
                ||   ||||||||||||||   || ||| |||| |||||| ||||||||||   |||
Db       9592 GACCGGCACCCCGTCTTCGCCCGGGCGCTCGACGAGATCTGCGCCCACCTCGACGGTCAC 9651

Qy        121 CTCGACCACCCCCTCCTCCCCCTCCTCACCCAAAACGACAACGACAACGAGGACGCGGCC 180
              ||||| |  ||||| |||  | |  |          ||    |||   |   |  | ||||
Db       9652 CTCGAACTGCCCCTGCTCGACGTGAT------GTTCGCGGCCGAGGGCAGCGCGGAGGCC 9705

Qy        181 GCACTGCTCCAGCAGACCCGCTACGCCCAGCCCGCCCTCTTCGCCTTCCAGGTCGCCCTC 240
              || |||||| |  |||| || ||| | |||  ||| || |||||| |  ||||||| |||
Db       9706 GCGCTGCTCGACGAGACGCGGTACACGCAGTGCGCGCTGTTCGCCCTGGAGGTCGCGCTC 9765

Qy        241 CACCGCCTCCTCACCGACGGCTACCACATCACCCCCCACTACTACGCCGGACACTCCCTC 300
                ||| ||| |   |||  |||   |||    ||  |       ||| |||||  ||
Db       9766 TTCCGGCTCGT---CGAGAGCTGGGGCATGCGGCCGGCCGCACTGCTCGGTCACTCGGTC 9822

Qy        301 GGCGAAATCACCGCCGCCCACCTCGCCGGCATCCTCACCCTCACCGACGCCACCACCCTC 360
              ||||| ||| ||||||| ||| |||||||  |  || | ||| |||||||| ||  |||
Db       9823 GGCGAGATCGCCGCCGCGCACGTCGCCGGTGTGTTCTCGCTCGCCGACGCCGCCCGCCTG 9882

Qy        361 ATCACCCAACGCGCCACCCTCATGCAAACCATGCCCCCC---GGCACCATGACCACCCTC 417
              || ||   ||| |  |||||||      ||||| ||   ||| | |||  |  || ||
Db       9883 GTCGCCGCGCGCGGCCGGCTCATGCAGGAGCTGCCCGCCGGTGGCGCGATGCTCGCCGTC 9942

Qy        418 CACACCACCCCCCACCACATCACCCACCACCTCACCGCCCACGAAAACGAC--------- 468
              ||  || |     || | |||  |     ||    |  | ||    ||
Db       9943 CAGGCCGCGGAGGACGAGATCCGCGTGTGGCTGGAGACGGAGGAGCGGTACGCGGGACGT 10002

Qy        469 CTCGCCATCGCCGCCATCAACACCCCCACCTCCCTCGTCATCAGCGGCACCCCCCACACC 528
              || | | |||||||| |||||  ||||     ||  |||| |   ||||  | |  || |
Db       10003 CTGGACGTCGCCGCCGTCAACGGCCCCGAGGCCGCCGTCCTGTCCGGCGACGCGGACGCG 10062

Qy        529 GTCCAACACATCACCACCCTCTGCCAACAACAAGGCATCAAAACCAAAACCCTCCCCACC 588
              |  |  |          |  |||          | ||| ||  |||    | || |    |
Db       10063 GCGCGGGAGGCGGAGGCGTACTGGTCCGGGCTCGGCCGCAGGACCCGCGCGCTGCGGGTC 10122

Qy        589 AACCACGCCTTCCACTCCCCCCACACCAACCCCATCCTCAACCAACTCCACCAGCACACC 648
              | |||||||||||||||| | ||||   ||  ||| ||| ||    |||  |      |
Db       10123 AGCCACGCCTTCCACTCCGCGCACATGGACGGCATGCTCGACGGGTTCCGCGCCGTCCTG 10182

Qy        649 CAAACCCTCACCTACCACCCACCCCACACCCCCCTCATCACCGCCAACA----------- 697
              | ||  |    | ||  |  |||  |    || |  || |    |  ||
Db       10183 GAGACGGTGGAGTTCCGGCGCCCCTCCCTGACCGTGGTCTCGAACGTCACCGGCCTGGCC 10242

Qy        698 -CCCCACCCGACCAACTCCTCACCCCCCACTACTGGACCCAACAAGCCCGCAACACCGTC 756
               ||   || ||| | ||   |  |||| | ||||||  ||  || | ||||  |||||||
Db       10243 GCCGGCCCGGACGACCTGTGCGACCCCGAGTACTGGGTCCGGCACGTCCGCGGCACCGTC 10302

Qy        757 GACTACGCCACCACCACCCAAACCCTCCACCAACACGGCGTCACCACCTACATCGAACTC 816
              || |  |  |  |  ||    ||| | | | | ||||||    |||| | | || ||
Db       10303 CGCTTCCTCGACGGCGTCCGTGTCCTGCGCGACCTCGGCGTGCGGACCTGCCTGGAGCTG 10362

Qy        817 GGACCCGACAACACCCTCACCACCCTCACCCACCACAACCTCCCCAACCCCCCCACCACC 876
              || ||||||      ||||||| || |  |   | ||  |||| |  || ||||| |
Db       10363 GGCCCCGACGGGGTCCTCACCGCCATGGCGGCCGACGGCCTCGCGGACACCCCCGCGGAT 10422

Qy        877 ACCCTCACCCTCACCCACCCCCACCACCACCCC 909
                ||    ||  | |||  |  |  |  | | ||

Db       10423 TCCGCTGCCGGCTCCCCCGTCGGCTCTCCCGCC 10455


RESULT 13
US-09-860-846-5
; Sequence 5, Application US/09860846
; Patent No. US20020164742A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/860,846
;  CURRENT FILING DATE:  2001-05-18
;  PRIOR APPLICATION NUMBER: 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 5
;   LENGTH: 36778
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-860-846-5

  Query Match             23.8%;  Score 220.2;  DB 9;  Length 36778;
  Best Local Similarity   56.5%;  Pred. No. 1.6e-38;
  Matches  527;  Conservative    0;  Mismatches  373;  Indels   33;  Gaps    5;

Qy          1 ACCGCATTCATCTGCTCCGGACAGGGCACCCAACGCCCCGGCATGGCCCACGGCCTCTAC 60
              ||||| ||| ||| | |||| |||||||  || |||||||||   | ||  |  || |||
Db       11273 ACCGCGTTCCTCTTCACCGGCCAGGGCAGTCAGCGCCCCGGCGCCGGCCGTGAGCTGTAC 11332

Qy         61 CACACCCACCCCGTCTTCGCCGCCGCACTCAACGACATCTGCACCCACCTCGACCCCCAC 120
                ||   ||||||||||||||   || ||| |||| |||||| ||||||||||   |||
Db       11333 GACCGGCACCCCGTCTTCGCCCGGGCGCTCGACGAGATCTGCGCCCACCTCGACGGTCAC 11392

Qy        121 CTCGACCACCCCCTCCTCCCCCTCCTCACCCAAAACGACAACGACAACGAGGACGCGGCC 180
              ||||| |  ||||| |||  | |  |          ||    |||   |   |  | ||||
Db       11393 CTCGAACTGCCCCTGCTCGACGTGAT------GTTCGCGGCCGAGGGCAGCGCGGAGGCC 11446

Qy        181 GCACTGCTCCAGCAGACCCGCTACGCCCAGCCCGCCCTCTTCGCCTTCCAGGTCGCCCTC 240
              || |||||| |  |||| || ||| | |||  ||| || |||||| |  ||||||| |||
Db       11447 GCGCTGCTCGACGAGACGCGGTACACGCAGTGCGCGCTGTTCGCCCTGGAGGTCGCGCTC 11506

Qy        241 CACCGCCTCCTCACCGACGGCTACCACATCACCCCCCACTACTACGCCGGACACTCCCTC 300
                ||| ||| |   |||  |||   |||    ||  |       ||| |||||  ||
Db       11507 TTCCGGCTCGT---CGAGAGCTGGGGCATGCGGCCGGCCGCACTGCTCGGTCACTCGGTC 11563

Qy        301 GGCGAAATCACCGCCGCCCACCTCGCCGGCATCCTCACCCTCACCGACGCCACCACCCTC 360
              ||||| ||| ||||||| ||| |||||||  |  || | ||| |||||||| ||  |||
Db       11564 GGCGAGATCGCCGCCGCGCACGTCGCCGGTGTGTTCTCGCTCGCCGACGCCGCCCGCCTG 11623

Qy        361 ATCACCCAACGCGCCACCCTCATGCAAACCATGCCCCCC---GGCACCATGACCACCCTC 417
              || ||   ||| |  |||||||      ||||| ||   ||| | |||  |  || ||
Db       11624 GTCGCCGCGCGCGGCCGGCTCATGCAGGAGCTGCCCGCCGGTGGCGCGATGCTCGCCGTC 11683

Qy        418 CACACCACCCCCCACCACATCACCCACCACCTCACCGCCCACGAAAACGAC--------- 468
              ||  || |     || | |||  |     ||    |  | ||    ||
Db       11684 CAGGCCGCGGAGGACGAGATCCGCGTGTGGCTGGAGACGGAGGAGCGGTACGCGGGACGT 11743

Qy        469 CTCGCCATCGCCGCCATCAACACCCCCACCTCCCTCGTCATCAGCGGCACCCCCCACACC 528
              || | | |||||||| |||||  ||||     ||  |||| |   ||||  | |  || |
Db       11744 CTGGACGTCGCCGCCGTCAACGGCCCCGAGGCCGCCGTCCTGTCCGGCGACGCGGACGCG 11803

Qy        529 GTCCAACACATCACCACCCTCTGCCAACAACAAGGCATCAAAACCAAAACCCTCCCCACC 588
              |  |  |          |  |||          | ||| ||  |||    | || |    |
Db       11804 GCGCGGGAGGCGGAGGCGTACTGGTCCGGGCTCGGCCGCAGGACCCGCGCGCTGCGGGTC 11863

Qy        589 AACCACGCCTTCCACTCCCCCCACACCAACCCCATCCTCAACCAACTCCACCAGCACACC 648
              | |||||||||||||||| | ||||   ||  ||| ||| ||    |||  |      |
Db       11864 AGCCACGCCTTCCACTCCGCGCACATGGACGGCATGCTCGACGGGTTCCGCGCCGTCCTG 11923
```

```
Qy        649 CAAACCCTCACCTACCACCCACCCCACACCCCCCTCATCACCGCCAACA---------- 697
              | ||| |   | ||  |  ||| |    || |  || |   |  ||
Db      11924 GAGACGGTGGAGTTCCGGCGCCCCTCCCTGACCGTGGTCTCGAACGTCACCGGCCTGGCC 11983

Qy        698 -CCCCACCCGACCAACTCCTCACCCCCCACTACTGGACCCAACAAGCCCGCAACACCGTC 756
              ||   || ||| | ||   |  |||| | ||||||  ||  || | ||||  |||||||
Db      11984 GCCGGCCCGGACGACCTGTGCGACCCCGAGTACTGGGTCCGGCACGTCCGCGGCACCGTC 12043

Qy        757 GACTACGCCACCACCACCCAAACCCTCCACCAACACGGCGTCACCACCTACATCGAACTC 816
              || |  |  |  ||     ||| | | | ||||||    |||| | | || ||
Db      12044 CGCTTCCTCGACGGCGTCCGTGTCCTGCGCGACCTCGGCGTGCGGACCTGCCTGGAGCTG 12103

Qy        817 GGACCCGACAACACCCTCACCACCCTCACCCACCACAACCTCCCCAACCCCCCCCACCACC 876
              || ||||||     ||||||| || |  |   | ||  |||| |  || ||||| |
Db      12104 GGCCCCGACGGGGTCCTCACCGCCATGGCGGCCGACGGCCTCGCGGACACCCCCGCGGAT 12163

Qy        877 ACCCTCACCCTCACCCACCCCCACCACCACCCC 909
              ||    ||  | ||| |  |  |   | | ||
Db      12164 TCCGCTGCCGGCTCCCCCGTCGGCTCTCCCGCC 12196


RESULT 14
US-09-836-821-5
; Sequence 5, Application US/09836821
; Publication No. US20030087405A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/836,821
;  CURRENT FILING DATE:  2001-04-17
;  PRIOR APPLICATION NUMBER: 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 5
;   LENGTH: 36778
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-836-821-5

  Query Match             23.8%;  Score 220.2;  DB 9;  Length 36778;
  Best Local Similarity   56.5%;  Pred. No. 1.6e-38;
  Matches  527;  Conservative    0;  Mismatches  373;  Indels   33;  Gaps    5;

Qy          1 ACCGCATTCATCTGCTCCGGACAGGGCACCCAACGCCCCGGCATGGCCCACGGCCTCTAC 60
              ||||| ||| ||| | |||| |||||||  || |||||||||   | ||  |  || |||
Db      11273 ACCGCGTTCCTCTTCACCGGCCAGGGCAGTCAGCGCCCCGGCGCCGGCCGTGAGCTGTAC 11332

Qy         61 CACACCCACCCCGTCTTCGCCGCCGCACTCAACGACATCTGCACCCACCTCGACCCCCAC 120
              ||   |||||||||||||||   || ||| |||| |||||| ||||||||||    |||
Db      11333 GACCGGCACCCCGTCTTCGCCCGGGCGCTCGACGAGATCTGCGCCCACCTCGACGGTCAC 11392

Qy        121 CTCGACCACCCCCTCCTCCCCCTCCTCACCCAAAACGACAACGACAACGAGGACGCGGCC 180
              ||||| |  ||||| |||  | ||  |          ||    |||   |   |  | ||||
Db      11393 CTCGAACTGCCCCTGCTCGACGTGAT------GTTCGCGGCCGAGGGCAGCGCGGAGGCC 11446

Qy        181 GCACTGCTCCAGCAGACCCGCTACGCCCAGCCCGCCCTCTTCGCCTTCCAGGTCGCCCTC 240
              || |||||| |  |||| || ||| | |||  ||| || |||||| |  ||||||| |||
Db      11447 GCGCTGCTCGACGAGACGCGGTACACGCAGTGCGCGCTGTTCGCCCTGGAGGTCGCGCTC 11506

Qy        241 CACCGCCTCCTCACCGACGGCTACCACATCACCCCCCACTACTACGCCGGACACTCCCTC 300
              |||  ||| |   |||  |||     |||     ||   |         ||| |||||  ||
Db      11507 TTCCGGCTCGT---CGAGAGCTGGGGCATGCGGCCGGCCGCACTGCTCGGTCACTCGGTC 11563

Qy        301 GGCGAAATCACCGCCGCCCACCTCGCCGGCATCCTCACCCTCACCGACGCCACCACCCTC 360
              ||||| ||| ||||||| ||| |||||||  |  || | ||| |||||||| ||  |||
Db      11564 GGCGAGATCGCCGCCGCGCACGTCGCCGGTGTGTTCTCGCTCGCCGACGCCGCCCGCCTG 11623

Qy        361 ATCACCCAACGCGCCACCCTCATGCAAACCATGCCCCCC---GGCACCATGACCACCCTC 417
              || ||   |||| |   |||||||||      ||||| ||   ||| | |||  | || ||
Db      11624 GTCGCCGCGCGCGGCCGGCTCATGCAGGAGCTGCCCGCCGGTGGCGCGATGCTCGCCGTC 11683

Qy        418 CACACCACCCCCCACCACATCACCCACCACCTCACCGCCCACGAAAACGAC--------- 468
              ||  || |    || | |||  |    ||     |  | ||     ||
Db      11684 CAGGCCGCGGAGGACGAGATCCGCGTGTGGCTGGAGACGGAGGAGCGGTACGCGGGACGT 11743

Qy        469 CTCGCCATCGCCGCCATCAACACCCCCACCTCCCTCGTCATCAGCGGCACCCCCCACACC 528
              || | | |||||||| ||||  ||||    ||  |||| |   ||||  | |  || ||
Db      11744 CTGGACGTCGCCGCCGTCAACGGCCCCGAGGCCGCCGTCCTGTCCGGCGACGCGGACGCG 11803

Qy        529 GTCCAACACATCACCACCCTCTGCCAACAACAAGGCATCAAAACCAAAACCCTCCCCACC 588
              | | |          |   |||        | |||  ||  |||    | || |    |
Db      11804 GCGCGGGAGGCGGAGGCGTACTGGTCCGGGCTCGGCCGCAGGACCCGCGCGCTGCGGGTC 11863

Qy        589 AACCACGCCTTCCACTCCCCCCACACCAACCCCATCCTCAACCAACTCCACCAGCACACC 648
              | |||||||||||||||| | ||||   ||  ||| ||| ||    ||| |   |
Db      11864 AGCCACGCCTTCCACTCCGCGCACATGGACGGCATGCTCGACGGGTTCCGCGCCGTCCTG 11923

Qy        649 CAAACCCTCACCTACCACCCACCCCACACCCCCCTCATCACCGCCAACA---------- 697
              | ||  |   | ||  |  ||| |    || |  || |   |  ||
Db      11924 GAGACGGTGGAGTTCCGGCGCCCCTCCCTGACCGTGGTCTCGAACGTCACCGGCCTGGCC 11983

Qy        698 -CCCCACCCGACCAACTCCTCACCCCCCACTACTGGACCCAACAAGCCCGCAACACCGTC 756
              ||   || ||| | ||   |  |||| | ||||||  ||  || | ||||  |||||||
Db      11984 GCCGGCCCGGACGACCTGTGCGACCCCGAGTACTGGGTCCGGCACGTCCGCGGCACCGTC 12043

Qy        757 GACTACGCCACCACCACCCAAACCCTCCACCAACACGGCGTCACCACCTACATCGAACTC 816
              || |  |  |  ||     ||| | | | ||||||    |||| | | || ||
Db      12044 CGCTTCCTCGACGGCGTCCGTGTCCTGCGCGACCTCGGCGTGCGGACCTGCCTGGAGCTG 12103

Qy        817 GGACCCGACAACACCCTCACCACCCTCACCCACCACAACCTCCCCAACCCCCCCACCACC 876
              || ||||||     ||||||| || |  |   | ||  |||| |  || ||||| |
Db      12104 GGCCCCGACGGGGTCCTCACCGCCATGGCGGCCGACGGCCTCGCGGACACCCCCGCGGAT 12163

Qy        877 ACCCTCACCCTCACCCACCCCCACCACCACCCC 909
              ||    ||  | ||| |  |  |   | | ||
Db      12164 TCCGCTGCCGGCTCCCCCGTCGGCTCTCCCGCC 12196


RESULT 15
US-09-861-289-5
; Sequence 5, Application US/09861289
; Patent No. US20020110897A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/861,289
;  CURRENT FILING DATE:  2001-05-18
;  PRIOR APPLICATION NUMBER: 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 5
;   LENGTH: 36778
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-861-289-5

  Query Match             23.8%;  Score 220.2;  DB 10;  Length 36778;
  Best Local Similarity   56.5%;  Pred. No. 1.6e-38;
  Matches  527;  Conservative    0;  Mismatches  373;  Indels   33;  Gaps    5;

Qy          1 ACCGCATTCATCTGCTCCGGACAGGGCACCCAACGCCCCGGCATGGCCCACGGCCTCTAC 60
              ||||| ||| ||| | |||| |||||||  || |||||||||   | ||  |  || |||
Db      11273 ACCGCGTTCCTCTTCACCGGCCAGGGCAGTCAGCGCCCCGGCGCCGGCCGTGAGCTGTAC 11332

Qy         61 CACACCCACCCCGTCTTCGCCGCCGCACTCAACGACATCTGCACCCACCTCGACCCCCAC 120
```

```
Db      11333 GACCGGCACCCCGTCTTCGCCCGGGCGCTCGACGAGATCTGCGCCCACCTCGACGGTCAC 11392

Qy        121 CTCGACCACCCCCTCCTCCCCCTCCTCACCCAAAACGACAACGACAACGAGGACGCGGCC 180

Db      11393 CTCGAACTGCCCCTGCTCGACGTGAT------GTTCGCGGCCGAGGGCAGCGCGGAGGCC 11446

Qy        181 GCACTGCTCCAGCAGACCCGCTACGCCCAGCCCGCCCTCTTCGCCTTCCAGGTCGCCCTC 240

Db      11447 GCGCTGCTCGACGAGACGCGGTACACGCAGTGCGCGCTGTTCGCCCTGGAGGTCGCGCTC 11506

Qy        241 CACCGCCTCCTCACCGACGGCTACCACATCACCCCCCACTACTACGCCGGACACTCCCTC 300

Db      11507 TTCCGGCTCGT---CGAGAGCTGGGGCATGCGGCCGGCCGCACTGCTCGGTCACTCGGTC 11563

Qy        301 GGCGAAATCACCGCCGCCCACCTCGCCGGCATCCTCACCCTCACCGACGCCACCACCCTC 360

Db      11564 GGCGAGATCGCCGCCGCGCACGTCGCCGGTGTGTTCTCGCTCGCCGACGCCGCCCGCCTG 11623

Qy        361 ATCACCCAACGCGCCACCCTCATGCAAACCATGCCCCCC---GGCACCATGACCACCCTC 417

Db      11624 GTCGCCGCGCGCGGCCGGCTCATGCAGGAGCTGCCCGCCGGTGGCGCGATGCTCGCCGTC 11683

Qy        418 CACACCACCCCCCACCACATCACCCACCACCTCACCGCCCACGAAAACGAC--------- 468

Db      11684 CAGGCCGCGGAGGACGAGATCCGCGTGTGGCTGGAGACGGAGGAGCGGTACGCGGGACGT 11743

Qy        469 CTCGCCATCGCCGCCATCAACACCCCCACCTCCCTCGTCATCAGCGGCACCCCCCACACC 528

Db      11744 CTGGACGTCGCCGCCGTCAACGGCCCCGAGGCCGCCGTCCTGTCCGGCGACGCGGACGCG 11803

Qy        529 GTCCAACACATCACCACCCTCTGCCAACAACAAGGCATCAAAACCAAAACCCTCCCCACC 588

Db      11804 GCGCGGGAGGCGGAGGCGTACTGGTCCGGGCTCGGCCGCAGGACCCGCGCGCTGCGGGTC 11863

Qy        589 AACCACGCCTTCCACTCCCCCCACACCAACCCCATCCTCAACCAACTCCACCAGCACACC 648

Db      11864 AGCCACGCCTTCCACTCCGCGCACATGGACGGCATGCTCGACGGGTTCCGCGCCGTCCTG 11923

Qy        649 CAAACCCTCACCTACCACCCACCCCACACCCCCCTCATCACCGCCAACA---------- 697

Db      11924 GAGACGGTGGAGTTCCGGCGCCCCTCCCTGACCGTGGTCTCGAACGTCACCGGCCTGGCC 11983

Qy        698 -CCCCACCCGACCAACTCCTCACCCCCCACTACTGGACCCAACAAGCCCGCAACACCGTC 756

Db      11984 GCCGGCCCGGACGACCTGTGCGACCCCGAGTACTGGGTCCGGCACGTCCGCGGCACCGTC 12043

Qy        757 GACTACGCCACCACCACCCAAACCCTCCACCAACACGGCGTCACCACCTACATCGAACTC 816

Db      12044 CGCTTCCTCGACGGCGTCCGTGTCCTGCGCGACCTCGGCGTGCGGACCTGCCTGGAGCTG 12103

Qy        817 GGACCCGACAACACCCTCACCACCCTCACCCACCACAACCTCCCCAACCCCCCCACCACC 876

Db      12104 GGCCCCGACGGGGTCCTCACCGCCATGGCGGCCGACGGCCTCGCGGACACCCCCGCGGAT 12163

Qy        877 ACCCTCACCCTCACCCACCCCCACCACCACCCC 909

Db      12164 TCCGCTGCCGGCTCCCCCGTCGGCTCTCCCGCC 12196

Search completed: June 18, 2003, 05:08:40
Job time : 115.288 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 17, 2003, 11:26:45 ; Search time 1184.48 Seconds
(without alignments)
12633.968 Million cell updates/sec

Title: US-09-914-286-1_COPY_7906_8829
Perfect score: 924
Sequence: 1 accgcattcatctgctccgg.....acccccaaacccacctcctc 924

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
c	1	256.4	27.7	1015	13	BI416650	BI416650 hasp001xm
c	2	254.8	27.6	931	13	BI416575	BI416575 hasp001xg
c	3	245.8	26.6	1197	13	BI416470	BI416470 hasp001xa
c	4	241	26.1	1005	17	BH148610	BH148610 ENTPZ67TR
c	5	237.4	25.7	1184	14	BM911719	BM911719 AGENCOURT
c	6	237	25.6	873	12	BG786249	BG786249 SEAUMC006
c	7	237	25.6	1138	17	AG152606	AG152606 Pan trogl
c	8	236.4	25.6	1082	10	BE455154	BE455154 HVSMEh009
c	9	236.2	25.6	1224	13	BI416537	BI416537 hasp001xe
	10	235	25.4	1161	14	BM912505	BM912505 AGENCOURT
c	11	232.8	25.2	1223	13	BI416573	BI416573 hasp001xg
c	12	230	24.9	1036	12	BF254665	BF254665 HVSMEf000
c	13	229.2	24.8	1196	12	BF256755	BF256755 HVSMEf001
c	14	228.4	24.7	1222	17	BH146899	BH146899 ENTQM47TF
c	15	225	24.4	1161	12	BF256588	BF256588 HVSMEf001
c	16	224.4	24.3	1042	13	BI416535	BI416535 hasp001xe
c	17	223.2	24.2	1251	10	BE455089	BE455089 HVSMEh009
c	18	223	24.1	1166	12	BF256751	BF256751 HVSMEf001
	19	222.8	24.1	1405	14	BQ798843	BQ798843 EST 1012
c	20	222.4	24.1	1138	12	BF256506	BF256506 HVSMEf001
c	21	222	24.0	1076	17	AG032994	AG032994 Pan trogl
c	22	221.8	24.0	1201	14	BQ063878	BQ063878 AGENCOURT
	23	221.4	24.0	960	14	BQ429353	BQ429353 AGENCOURT
c	24	221.2	23.9	937	13	BI416476	BI416476 hasp001xa
c	25	220.6	23.9	1004	17	AG134371	AG134371 Pan trogl
c	26	220.6	23.9	1189	10	BE455091	BE455091 HVSMEh009
c	27	219.8	23.8	877	13	BI416653	BI416653 hasp001xm
c	28	219.8	23.8	1116	12	BF256617	BF256617 HVSMEf001
c	29	219.4	23.7	1080	13	BI416623	BI416623 hasp001xk
	30	219.2	23.7	1303	14	BM908878	BM908878 AGENCOURT
c	31	219	23.7	1186	12	BF256757	BF256757 HVSMEf001
c	32	218	23.6	924	10	BE455162	BE455162 HVSMEh009
c	33	217.2	23.5	946	17	AG080623	AG080623 Pan trogl
c	34	217	23.5	1086	12	BG809658	BG809658 mgct001xe
c	35	216.8	23.5	1120	14	BQ686415	BQ686415 AGENCOURT
c	36	216.6	23.4	1037	17	AG146864	AG146864 Pan trogl
c	37	216.6	23.4	1137	17	AG078502	AG078502 Pan trogl
c	38	216.4	23.4	1089	14	BQ930965	BQ930965 AGENCOURT
c	39	216	23.4	1088	12	BF256580	BF256580 HVSMEf001
c	40	215.4	23.3	1177	17	AG081646	AG081646 Pan trogl
c	41	215	23.3	825	12	BG786255	BG786255 SEAUMC006
c	42	215	23.3	1007	17	AG134493	AG134493 Pan trogl
c	43	215	23.3	1152	12	BF256631	BF256631 HVSMEf001
c	44	214.8	23.2	883	17	AZ186646	AZ186646 SP_1007_B
	45	214.6	23.2	1268	13	BI752215	BI752215 603022405

ALIGNMENTS

RESULT 1
BI416650/c
LOCUS BI416650 1015 bp mRNA linear EST 15-AUG-2001
DEFINITION hasp001xm01f Heterobasidion annosum - Scots pine infection stage subtraction cDNA library (hasp) Pinus sylvestris/Heterobasidion annosum cDNA clone hasp001xm01f, mRNA sequence.
ACCESSION BI416650
VERSION BI416650.1 GI:15187673
KEYWORDS EST.
SOURCE Pinus sylvestris/Heterobasidion annosum.
ORGANISM Pinus sylvestris/Heterobasidion annosum
Eukaryota; mixed EST libraries.
REFERENCE 1 (bases 1 to 1015)
AUTHORS Asiegbu,F.O., Nahalkova,J., Choi,W.; Stenlid,J. and Dean,R.A.
TITLE Expressed sequence tags of randomly selected cDNA clones from the interaction of the root rot fungus (Heterobasidion annosum) with seedling roots of Scots pine (Pinus sylvestris)
JOURNAL Unpublished (2001)
COMMENT Contact: Fred O. Asiegbu
Dept. of Forest Mycology & Pathology
Swedish University of Agriculture, Box 7026,S-750 07, Uppsala, Sweden
Tel: +46 18 67 15 98
Fax: +46 18 30 92 45
Email: Fred.Asiegbu@mykopat.slu.se
Seq primer: T7 primer.
FEATURES Location/Qualifiers
source 1. .1015
/organism="Pinus sylvestris/Heterobasidion annosum"

/db_xref="taxon:169015"
/clone="hasp001xm01f"
/clone_lib="Heterobasidion annosum - Scots pine infection stage subtraction cDNA library (hasp)"
/dev_stage="Seedling roots of scots pine were infected for 6 days with H. annosum"
/note="Vector: pT-Adv; Site_1: EcoRI; The subtractive hybridization cDNA library was constructed from scots pine roots infected for 6-days with mycelia of Heterobasidon annosum (FP5)."

BASE COUNT 21 a 6 c 961 g 16 t 11 others
ORIGIN

Query Match 27.7%; Score 256.4; DB 13; Length 1015;
Best Local Similarity 55.7%; Pred. No. 5.5e-32;
Matches 500; Conservative 0; Mismatches 396; Indels 2; Gaps 1;

```
Qy         27 CACCCAACGCCCCGGCATGGCCCACGGCCTCTACCACACCCACCCCGTCTTCGCCGCCGC 86
Db        926 CACCCCCCTCCCCCCCCCCCCCCCCCCCCCCCCCCCCCNCCCCCTCCCCCCCCCCCCCCCC 867

Qy         87 ACTCAACGACATCTGCACCCACCTCGACCCCCACCTCGACCACCCCCTCCTCCCCCTCCT 146
Db        866 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 807

Qy        147 CACCCAAAACGACAACGACAACGAGGACGCGGCCGCACTGCTCCAGCAGACCCGCTACGC 206
Db        806 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 747

Qy        207 CCAGCCCGCCCTCTTCGCCTTCCAGGTCGCCCTCCACCGCCTCCTCACCGACGGCTACCA 266
Db        746 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCACCCCCCCCCCCCCCCCCCCCCC 687

Qy        267 CATCACCCCCCACTACTACGCCGGACACTCCCTCGGCGAAATCACCGCCGCCCACCTCGC 326
Db        686 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 627

Qy        327 CGGCATCCTCACCCTCACCGACGCCACCACCCTCATCACCCAACGCGCCACCCTCATGCA 386
Db        626 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCTCCCCCCCCCCCCCCCCCCC 567

Qy        387 AACCATGCCCCCCGGCACCATGACCACCCTCCACACCACCCCCCCACCACATCACCCACCA 446
Db        566 TCCCACCCCCCCCCCCCCCCCCCCCCCCCCCCTACCCCCCCCCCCACCCCCCCCCCCCCC 507

Qy        447 CCTCACCGCCCACGAAAACGACCTCGCCATCGCCGCCATCAACACCCCCACCTCCCTCGT 506
Db        506 CCCCCCCCCCCCCCCCCCCCACCCCCACCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 447

Qy        507 CATCAGCGGCACCCCCCCACACCGTCCAACACATCACCACCCTCTGCCAACAACAAGGCAT 566
Db        446 CCCCCCCCCCCACCCCCCCCCCCCCCCCCTCCCCCCCCCCCCCCCCCCCCTCCCCCCCCCC 387

Qy        567 CAAAACCAAAACCCTCCCCACCAACCACGCCTTCCACTCCCCCCACACCAACCCCATCCT 626
Db        386 CCCCCCCCCACCCCCCCCCCCC--CCCCCCCTCCCCNCCCCCCCCCCCCCCCCCCCCCCC 329

Qy        627 CAACCAACTCCACCAGCACACCCAAACCCTCACCTACCACCCACCCCACACCCCCCTCAT 686
Db        328 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCNCCCCCCTCCCCTCCCCC 269

Qy        687 CACCGCCAACACCCCACCCGACCAACTCCTCACCCCCCACTACTGGACCCAACAAGCCCG 746
Db        268 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCGCCCCCCNCCCCCC 209

Qy        747 CAACACCGTCGACTACGCCACCACCACCCAAACCCTCCACCAACACGGCGTCACCACCTA 806
Db        208 CCCCCACCCCNCCCCCNCCCCCCCCCCTCCCCCCCCCCCCCCGTCCCCCCCCCCCCACGCC 149

Qy        807 CATCGAACTCGGACCCGACAACACCCTCACCACCCTCACCCACCACAACCTCCCCAACCC 866
Db        148 CCCCCCCNCCGTCCCCCCCCCCCAGTAGCCCCCCCCCCCCCCCCCCCCACCCCCCCCACCCC 89
```

```
Qy        867 CCCCACCACCACCCTCACCCTCACCCACCCCCACCACCACCCCCAAACCCACCTCCTC 924
Db         88 CCCCCCCCCCCCCCCNCCCCCCCCCCCCCCCCCCCNCCCCCCCCCCNCCCCCCCCCCCCCC 31
```

RESULT 2
BI416575/c
LOCUS BI416575 931 bp mRNA linear EST 15-AUG-2001
DEFINITION hasp001xg15f Heterobasidion annosum - Scots pine infection stage subtraction cDNA library (hasp) Pinus sylvestris/Heterobasidion annosum cDNA clone hasp001xg15f, mRNA sequence.
ACCESSION BI416575
VERSION BI416575.1 GI:15187598
KEYWORDS EST.
SOURCE Pinus sylvestris/Heterobasidion annosum.
ORGANISM Pinus sylvestris/Heterobasidion annosum
Eukaryota; mixed EST libraries.
REFERENCE 1 (bases 1 to 931)
AUTHORS Asiegbu,F.O., Nahalkova,J., Choi,W., Stenlid,J. and Dean,R.A.
TITLE Expressed sequence tags of randomly selected cDNA clones from the interaction of the root rot fungus (Heterobasidion annosum) with seedling roots of Scots pine (Pinus sylvestris)
JOURNAL Unpublished (2001)
COMMENT Contact: Fred O. Asiegbu
Dept. of Forest Mycology & Pathology
Swedish University of Agriculture, Box 7026,S-750 07, Uppsala, Sweden
Tel: +46 18 67 15 98
Fax: +46 18 30 92 45
Email: Fred.Asiegbu@mykopat.slu.se
Seq primer: T7 primer.
FEATURES Location/Qualifiers
source 1. .931
/organism="Pinus sylvestris/Heterobasidion annosum"
/db_xref="taxon:169015"
/clone="hasp001xg15f"
/clone_lib="Heterobasidion annosum - Scots pine infection stage subtraction cDNA library (hasp)"
/dev_stage="Seedling roots of scots pine were infected for 6 days with H. annosum"
/note="Vector: pT-Adv; Site_1: EcoRI; The subtractive hybridization cDNA library was constructed from scots pine roots infected for 6-days with mycelia of Heterobasidon annosum (FP5)."
BASE COUNT 9 a 7 c 884 g 15 t 16 others
ORIGIN

Query Match 27.6%; Score 254.8; DB 13; Length 931;
Best Local Similarity 54.6%; Pred. No. 9.9e-32;
Matches 490; Conservative 0; Mismatches 408; Indels 0; Gaps 0;

```
Qy         27 CACCCAACGCCCCGGCATGGCCCACGGCCTCTACCACACCCACCCCGTCTTCGCCGCCGC 86
Db        915 CCCCCCCCCCCCCCCCCCCCCCCCACCACACCCCCCCCCCCCCACCCCCCCCCCCCCCCC 856

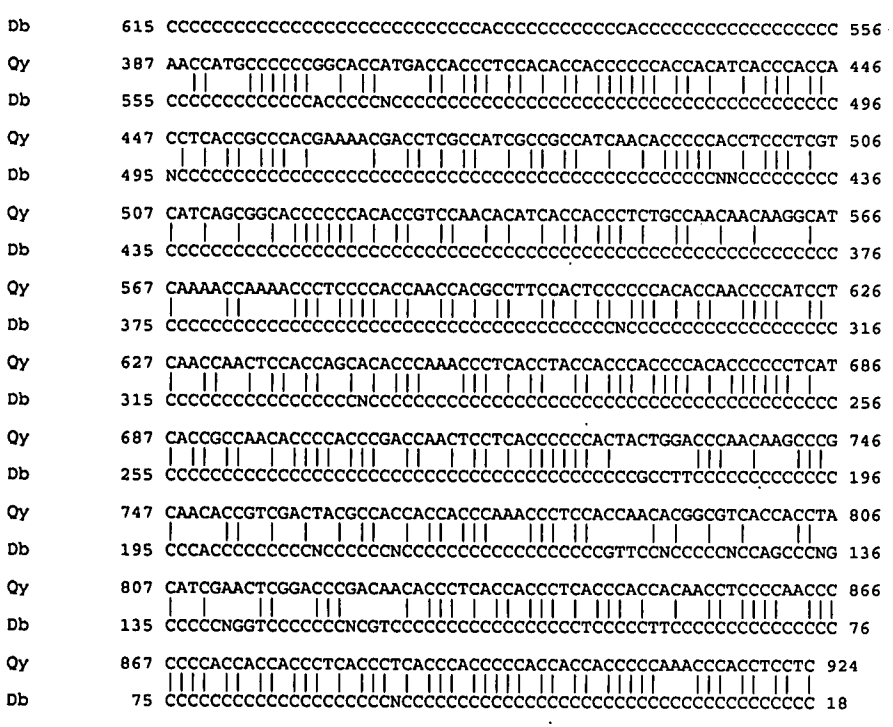
Qy         87 ACTCAACGACATCTGCACCCACCTCGACCCCCACCTCGACCACCCCCTCCTCCCCCTCCT 146
Db        855 CCCCCCCCCCCNCCCCCCCCCCCCCCCCCCCCCCCCCCCCCACCCCCCCCCCNCCCCCCCCC 796

Qy        147 CACCCAAAACGACAACGACAACGAGGACGCGGCCGCACTGCTCCAGCAGACCCGCTACGC 206
Db        795 CCCCCCACCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 736

Qy        207 CCAGCCCGCCCTCTTCGCCTTCCAGGTCGCCCTCCACCGCCTCCTCACCGACGGCTACCA 266
Db        735 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCACCCCCCCCCCCCCCCCCCCCCCCCCC 676

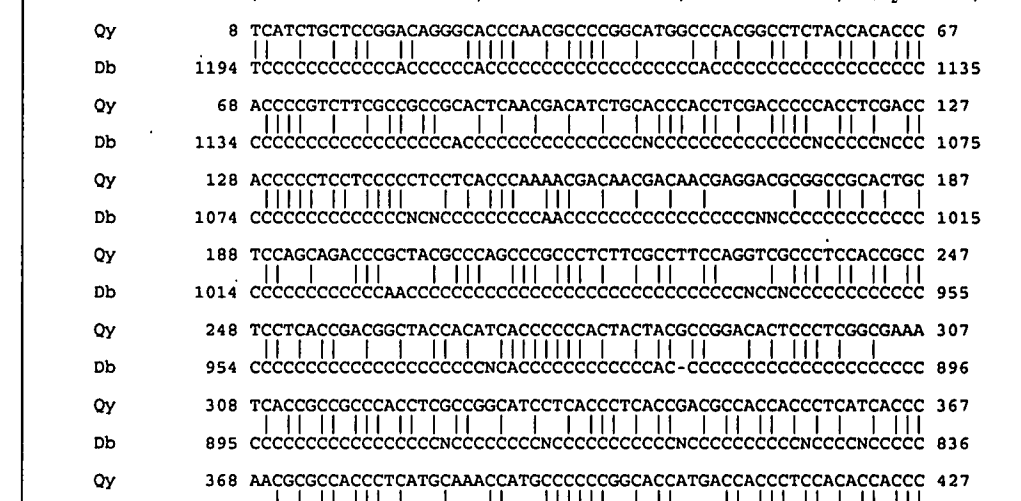
Qy        267 CATCACCCCCCACTACTACGCCGGACACTCCCTCGGCGAAATCACCGCCGCCCACCTCGC 326
Db        675 CCCCCCCCCCCCCCACCCCCCCCCCCCCCCCCCACCCCCCCCCCCCCCCCCCACCCCCCC 616

Qy        327 CGGCATCCTCACCCTCACCGACGCCACCACCCTCATCACCCAACGCGCCACCCTCATGCA 386
```



RESULT 3
BI416470/c
LOCUS BI416470 1197 bp mRNA linear EST 15-AUG-2001
DEFINITION hasp001xa01f Heterobasidion annosum - Scots pine infection stage subtraction cDNA library (hasp) Pinus sylvestris/Heterobasidion annosum cDNA clone hasp001xa01f, mRNA sequence.
ACCESSION BI416470
VERSION BI416470.1 GI:15187493
KEYWORDS EST.
SOURCE Pinus sylvestris/Heterobasidion annosum.
ORGANISM Pinus sylvestris/Heterobasidion annosum
Eukaryota; mixed EST libraries.
REFERENCE 1 (bases 1 to 1197)
AUTHORS Asiegbu,F.O., Nahalkova,J., Choi,W., Stenlid,J. and Dean,R.A.
TITLE Expressed sequence tags of randomly selected cDNA clones from the interaction of the root rot fungus (Heterobasidion annosum) with seedling roots of Scots pine (Pinus sylvestris)
JOURNAL Unpublished (2001)
COMMENT Contact: Fred O. Asiegbu
Dept. of Forest Mycology & Pathology
Swedish University of Agriculture, Box 7026,S-750 07, Uppsala, Sweden
Tel: +46 18 67 15 98
Fax: +46 18 30 92 45
Email: Fred.Asiegbu@mykopat.slu.se
Seq primer: T7 primer.
FEATURES Location/Qualifiers
source 1. .1197
/organism="Pinus sylvestris/Heterobasidion annosum"
/db_xref="taxon:169015"
/clone="hasp001xa01f"
/clone_lib="Heterobasidion annosum - Scots pine infection stage subtraction cDNA library (hasp)"
/dev_stage="Seedling roots of scots pine were infected for 6 days with H. annosum"
/note="Vector: pT-Adv; Site_1: EcoRI; The subtractive hybridization cDNA library was constructed from scots pine roots infected for 6-days with mycelia of Heterobasidon annosum (FP5)."
BASE COUNT 20 a 28 c 1036 g 28 t 85 others
ORIGIN

Query Match 26.6%; Score 245.8; DB 13; Length 1197;
Best Local Similarity 52.7%; Pred. No. 2.7e-30;
Matches 483; Conservative 0; Mismatches 433; Indels 1; Gaps 1;



```
Db        295 GGCCTCCCCNCNCCNNC 279


RESULT 4
BH148610/c
LOCUS       BH148610                1005 bp    DNA     linear   GSS 27-AUG-2001
DEFINITION  ENTPZ67TR Entamoeba histolytica Sheared DNA Entamoeba histolytica
            genomic, DNA sequence.
ACCESSION   BH148610
VERSION     BH148610.1  GI:15308098
KEYWORDS    GSS.
SOURCE      Entamoeba histolytica.
  ORGANISM  Entamoeba histolytica
            Eukaryota; Entamoebidae; Entamoeba.
REFERENCE   1  (bases 1 to 1005)
  AUTHORS   Loftus,B., Wang,Z., Van Aken,S. and Fraser,C.
  TITLE     Determination of clone end sequences from Entamoeba histolytica
            HM1:IMSS sheared DNA library (2001)
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Brendan J Loftus
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0208
            Fax: 301 838 3543
            Email: bjloftus@tigr.org
            Clones are derived from the Entamoeba histolytica HM1:IMSS sheared
            DNA library
            Seq primer: M13-Reverse
            Class: shotgun
            High quality sequence start: 501
            High quality sequence stop: 604.
FEATURES             Location/Qualifiers
     source          1. .1005
                     /organism="Entamoeba histolytica"
                     /strain="HM1:IMSS" .
                     /db_xref="taxon:5759"
                     /clone_lib="Entamoeba histolytica Sheared DNA"
                     /note="Vector: pHOS1; Site_1: Bst I; Constructed at The
                     Institute for Genomic Research (TIGR), Rockville, MD.
                     Genomic DNA isolated from broth cultures of E. histolytica
                     using a  method described by Clark and Diamond (Clark,
                     C.G., and Diamond, L.S.  (1993) Entamoeba histolytica: a
                     method for isolate identification.  Exp. Parasitol.
                     77:450.). The DNA was mechanically sheared to give a
                     tight size distribution (~2 kb). The v + i method used for
                     the library  construction is described in detail in Smith,
                     H.O. and Venter, J.C.  (Making small insert libraries for
                     whole genome shotgun sequencing  projects. In Genome
                     Sequencing: A Practical Approach, eds. M. Vaudin  and B.
                     Barell, Oxford University Press, 1999)."
BASE COUNT       24 a     13 c    838 g    130 t
ORIGIN

  Query Match            26.1%;  Score 241;  DB 17;  Length 1005;
  Best Local Similarity  55.6%;  Pred. No. 1.6e-29;
  Matches  484;  Conservative    0;  Mismatches  385;  Indels    2;  Gaps    1;

Qy         47 CCCACGGCCTCTACCACACCCACCCCGTCTTCGCCGCCGCACTCAACGACATCTGCACCC 106
Db       1004 CCCCCCCACACACCCCCCCACCCCCCCCCCCCTCCCCCCCCCCCCCCCCCCCCCCCCCC 945

Qy        107 ACCTCGACCCCCACCTCGACCACCCCCTCCTCCCCCTCCTCACCCAAAACGACAACGACA 166
Db        944 CCCCCCCCCACCCCCCCCCCCCCCCCCCCCACCCCCCCCCCCCCCCCCCCCCCACCACCCCC 885

Qy        167 ACGAGGACGCGGCCGCACTGCTCCAGCAGACCCGCTACGCCCAGCCCGCCCTCTTCGCCT 226
Db        884 CCACCCCCCCCACCCCCCCCCACCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 825

Qy        227 TCCAGGTCGCCCTCCACCGCCTCCTCACCGACGGCTACCACATCACCCCCCACTACTACG 286
```

```
Db        824 CACCCCCCCCCCCCCCCCCCCCCACCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 765

Qy        287 CCGGACACTCCCTCGGCGAAATCACCGCCGCCCACCTCGCCGGCATCCTCACCCTCACCG 346
Db        764 CCCCCCCCCACCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 705

Qy        347 ACGCCACCACCCTCATCACCCAACGCGCCACCCTCATGCAAACCATGCCCCCCGGCACCA 406
Db        704 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 645

Qy        407 TGACCACCCTCCACACCACCCCCCACCACATCACCCACCACCTCACCGCCCACGAAAACG 466
Db        644 CCCCCCCCCCCCACCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCGCCCCCCCCCCCC 585

Qy        467 ACCTCGCCATCGCCGCCATCAACACCCCCACCTCCCTCGTCATCAGCGGCACCCCCCACA 526
Db        584 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCACCCCCCCCCCCCCCCCCCC 525

Qy        527 CCGTCCAACACATCACCACCCTCTGCCAACAACAAGGCATCAAAACCAAAACCCTCCCCA 586
Db        524 CCCCCCCCCCCCCCCCCCCCCCCCCCCCAACACCCCCCAACCCCCCCCCCCCCCCACCCCC 465

Qy        587 CCAACCACGCCTTCCACTCCCCCCACACCAACCCCCATCCTCAACCAACTCCACCAGCACA 646
Db        464 CCCCCCCCCCCCACCCCCACACCCCCACCCCCCCCCCCCCCCCCACCCCCCCCCCCCCCCA 405

Qy        647 CCCAAACCCTCACCTACCACCCACCCCACACCCCCCTCATCACCGCCAACACCCCACCCG 706
Db        404 CCAACACCCCCCCTCCCCCCCCCCCCCCCACCCCCCCCCCCCCCCCCCCCACCCCCCCCCC 345

Qy        707 ACCAACTCCTCACCCCCCACTACTGGACCCAACAAGCCCGCAACACCGTCGACTACGCCA 766
Db        344 CCCCACCCCCCCCCCCCACCCCCCCCCCCCCCCCCCCACCCCCCCCCCCCCCCACCCCCCCA 285

Qy        767 CCACCACCC--AAACCCTCCACCAACACGGCGTCACCACCTACATCGAACTCGGACCCGA 824
Db        284 CCCCCCCCCCAAAACACCCCAAAAACCCCCCCACCCCCCCCCCCCCCCCCCCCACCCCAA 225

Qy        825 CAACACCCTCACCACCCTCACCCACCACAACCTCCCCAACCCCCCCACCACCACCCTCAC 884
Db        224 CCACCCAAAAACCCCAACCACCCCCCCCACCCCCCCCCGCGCCAAAACAACCACCCCCAC 165

Qy        885 CCTCACCCACCCCCACCACCACCCCCAAACC 915
Db        164 CAAAAAACCCCCCCACCAACCCACCCCCCCC 134


RESULT 5
BM911719/c
LOCUS       BM911719                1184 bp    mRNA    linear   EST 12-MAR-2002
DEFINITION  AGENCOURT_6640314 NIH_MGC_41 Homo sapiens cDNA clone IMAGE:5470806
            5', mRNA sequence.
ACCESSION   BM911719
VERSION     BM911719.1  GI:19362098
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1184)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: DCTD/DTP
             cDNA Library Preparation: Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1978  row: d  column: 07
```

```
          High quality sequence stop: 289.
FEATURES             Location/Qualifiers
     source          1. .1184
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5470806"
                     /clone_lib="NIH_MGC_41"
                     /tissue_type="amelanotic melanoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: skin; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming. Directionally cloned
                     into EcoRI/XhoI sites using the following 5' adaptor:
                     GGCACGAG(G). Library constructed by Ling Hong in the
                     laboratory of Gerald M. Rubin (University of California,
                     Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
                     Superscript II RT (Life Technologies). Note: this is a
                     NIH_MGC Library."
BASE COUNT       66 a     92 c    843 g     42 t    141 others
ORIGIN

  Query Match           25.7%;  Score 237.4;  DB 14;  Length 1184;
  Best Local Similarity   49.7%;  Pred. No. 6e-29;
  Matches  427;  Conservative    0;  Mismatches  432;  Indels    0;  Gaps    0;

Qy         66 CCACCCCGTCTTCGCCGCCGCACTCAACGACATCTGCACCCACCTCGACCCCCACCTCGA 125
              || ||||  |  | || || | | |  |  | |  | ||| ||    ||||| | |
Db       1183 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCNCCCCCCCCCCCCCCNCNCCCCCNCNNCCC 1124

Qy        126 CCACCCCCTCCTCCCCCTCCTCACCCAAAACGACAACGACAACGAGGACGCGGCCGCACT 185
              ||·|| ||    | | | || | |||        | |      |      | |  || | |
Db       1123 CCNCCTCCNNNNCNCNCCCCCCCCCCCCCCNNNCCCCCCNCCCCCCCCCCCCCCCNCCCN 1064

Qy        186 GCTCCAGCAGACCCGCTACGCCCAGCCCGCCCTCTTCGCCTTCCAGGTCGCCCTCCACCG 245
               | ||  |   ||| |  | ||  ||| ||      | ||  ||     |  | || ||
Db       1063 CCCCCCCCCCCCCCCNCCCCCNCCCCCCCNNNNCCCCCCCCCCNCCCCNNCNCCCCCN 1004

Qy        246 CCTCCTCACCGACGGCTACCACATCACCCCCCACTACTACGCCGGACACTCCCTCGGCGA 305
              || |  | ||  |  |  || |  | ||||||    |  | ||   | | ||  |
Db       1003 CCCCNNCCCCNCCCCCCCCCCCCCCCCCCCCCCCNNNNCCCCCCCCNNCCCNCCNCCCCNNN 944

Qy        306 AATCACCGCCGCCCACCTCGCCGGCATCCTCACCCTCACCGACGCCACCACCCTCATCAC 365
                | |  || ||| || | ||  |  ||     ||   ||  | || || ||| |  | |
Db        943 NCCCCCNNCCNCCCCCCCCCCCCCCCCCCCCNNNNCCNNNCCCCCCCCCCCCCCCCCCCCC 884

Qy        366 CCAACGCGCCACCCTCATGCAAACCATGCCCCCCGGCACCATGACCACCCTCCACACCAC 425
              ||  | | ||  ||      |  ||  ||||||  | ||    || ||| || |    |
Db        883 CCCCCCCCCCNNCCNNNCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCNNNC 824

Qy        426 CCCCCACCACATCACCCACCACCTCACCGCCCACGAAAACGACCTCGCCATCGCCGCCAT 485
              || || || |  | ||| || || | || | | |          || | |   | |  ||
Db        823 CCNCCCCCCCCCCCCCCCCCCCCCCCCCCCCNCCCCCCNNNNCCCCCCCNNCCCCNCCCCC 764

Qy        486 CAACACCCCCACCTCCCTCGTCATCAGCGGCACCCCCCACACCGTCCAACACATCACCAC 545
              |  |  |||| || ||| |      |  |  | || | |   ||  ||  |  | ||| |
Db        763 CCCCCNCCCCCCCCCCCCCCCNCCCCCCNCCCCCNCNCCNNCCCCCCCCCCCCCCCCCCC 704

Qy        546 CCTCTGCCAACAACAAGGCATCAAAACCAAAACCCTCCCCACCAACCACGCCTTCCACTC 605
              || |  ||  |  |       |  ||    ||   || ||  || | ||  || | | |
Db        703 CCCCCCCCCNCCNCCCCNNCCCCCCCCCCCCCCCCNNNNCCCCCCCCCCCCCCCCCCCCCC 644

Qy        606 CCCCCACACCAACCCCATCCTCAACCAACTCCACCAGCACACCCAAACCCTCACCTACCA 665
              |||||  |  |  || || |  | || | | ||  | | ||    ||| | ||  ||
Db        643 CCCCCCCCCNCCCNCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCNCCCCCCCCCCCCCCCC 584

Qy        666 CCCACCCCACACCCCCCTCATCACCGCCAACACCCCACCCGACCAACTCCTCACCCCCCA 725
              ||| |||| | |||||| | ·| || ||  | |||| |||  ||  | || |.||||||
Db        583 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCNCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 524

Qy        726 CTACTGGACCCAACAAGCCCGCAACACCGTCGACTACGCCACCACCACCCAAACCCTCCA 785
              |      | |     ||| |  | ||  |  |  | || || || |||   ||| ||
Db        523 CCCNNNCCCNCNNNNCCCCCCCCCCCCCCCCCCCCNCCCCCCCCCCCCCCCCCCCCCCCCC 464

Qy        786 CCAACACGGCGTCACCACCTACATCGAACTCGGACCCGACAACACCCTCACCACCCTCAC 845
```

```
              || | |  |  | || ||  |  |   | |   |||  |  | ||| | || ||| | |
Db        463 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCNCCCCCCCCCCCC 404

Qy        846 CCACCACAACCTCCCCAACCCCCCCACCACCACCCTCACCCTCACCCACCCCCACCACCA 905
                 || |  || ||||  |||||||       ||| | ||| | ||| | ||| || ||
Db        403 NNNCCCCCCCCCCCCCCCCCCCCCCCNNNCNNCCCCCCCCCCCCCCCCCCNCCCCCCCCCC 344

Qy        906 CCCCCAAACCCACCTCCTC 924
              |||||   | | || || |
Db        343 CCCCCNNCCNCNCCNCCNC 325


RESULT 6
BG786249/c
LOCUS       BG786249                 873 bp    mRNA    linear   EST 20-MAY-2001
DEFINITION  SEAUMC006206 Sea urchin primary mesenchyme cell cDNA library
            Strongylocentrotus purpuratus cDNA clone PC_0028_A2_A02_MR 5', mRNA
            sequence.
ACCESSION   BG786249
VERSION     BG786249.1  GI:14157262
KEYWORDS    EST.
SOURCE      Strongylocentrotus purpuratus.
  ORGANISM  Strongylocentrotus purpuratus
            Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
            Echinoidea; Euechinoidea; Echinacea; Echinoida;
            Strongylocentrotidae; Strongylocentrotus.
REFERENCE   1  (bases 1 to 873)
  AUTHORS   Zhu,X., Mahairas,G., Illies,M.R., Cameron,R.A., Davidson,E.H. and
            Ettensohn,C.A.
  TITLE     A large scale analysis of mRNAs expressed by primary mesenchyme
            cells of the sea urchin embryo
  JOURNAL   Development 128 (13), 2615-2627 (2001)
  MEDLINE   21384984
COMMENT     Contact: Ettensohn CA
            Dept. Biol. Sci.
            Carnegie Mellon University
            4400 Fifth Avenue, Pittsburgh, PA 15213, USA
            Tel: +1 412 268 5849
            Email: ettensohn@andrew.cmu.edu.
FEATURES             Location/Qualifiers
     source          1. .873
                     /organism="Strongylocentrotus purpuratus"
                     /db_xref="taxon:7668"
                     /clone="PC_0028_A2_A02_MR"
                     /clone_lib="Sea urchin primary mesenchyme cell cDNA
                     library"
                     /tissue_type="embryo"
                     /cell_type="primary mesenchyme cells"
                     /lab_host="E.coli"
                     /note="Vector: pSPORT1; Site_1: NotI; Site_2: SalI; oligo
                     dt priming from poly A+ RNA, directionally cloned"
BASE COUNT       10 a      0 c    740 g     85 t     38 others
ORIGIN

  Query Match           25.6%;  Score 237;  DB 12;  Length 873;
  Best Local Similarity   54.7%;  Pred. No. 7e-29;
  Matches  450;  Conservative    0;  Mismatches  371;  Indels    2;  Gaps    1;

Qy        102 CACCCACCTCGACCCCCACCTCGACCACCCCCTCCTCCCCCTCCTCACCCAAAACGACAA 161
              | ||| || |  |||||||| |  || | ||| ||  |||| || | |||     | |
Db        835 CCCCCCCCCCCCCCCCCACCCCCCCCCCACCCCCCCACCCCCCCCCCCCCCCCCANCCC 776

Qy        162 CGACAACGAGGACGCGGCCGCACTGCTCCAGCAGACCCGCTACGCCCAGCCCGCCCTCTT 221
              | ||  |    | |  || | |  | ||  |    || | || | |  ||| | |
Db        775 CCACCCCCCACCCCCCCCCCCCCCCCCCCCCCCCCCCNCCCCCACACACCCCCCCCACCACC 716

Qy        222 CGCCTTCCAGGTCGCCCTCCACCGCCTCCTCACCGACGGCTACCACATCACCCCCCACTA 281
              ||  ||     | ||| || || || || | ||  |   ||  | ||||  |||||
Db        715 ACCCCCCCCCACCCCCCCCCCCCCCCCCCCCCCCCNCCACCCCCCCAACCACCACCCACAC 656

Qy        282 CTACGCCGGACACTCCCTCGGCGAAATCACCGCCGCCCACCTCGCCGGCATCCTCACCCT 341
              | | |     |  || |  | |   ||| || ||| || | ||  |  || | |||
Db        655 CCCCCCAACCACCCACCCCCCCCCACCCCACNCCCNCCCCCCCCCCCCCCNCCACCCCCCCCC 596
```

```
Qy        342 CACCGACGCCACCACCCTCATCACCCAACGCGCCACCCTCATGCAAACCATGCCCCCCGG 401
              | ||  | |||||  ||  |  | |||  | | || ||| |   | | ||   ||||||
Db        595 CNCCCNCACCACCCACCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCACCCCCCCCCCCCC 536

Qy        402 CACCATGACCACCCTCCACACCACCCCCCCACCACATCACCCACCACCTCACCGCCCACGA 461
                ||   ||| ||| |||| || |||||||| |  | ||||| |    || ||| |
Db        535 ACCCCCCACCCCCCCCCCACCCCCCCCCCCACCCCCCCCCCCCACCCCACNCCCCCCCCCCC 476

Qy        462 AAACGACCTCGCCATCGCCGCCATCAACACCCCCACCTCCCTCGTCATCAGCGGCACCCC 521
                ||| | |    |  | |||  | | ||| | || | | |  |  |      | |  |
Db        475 CCCAAACCCCCCNCNCCNCNCCACNACCNCCCNCCCCNCACCCCCCCCCCCCANCCCCNNC 416

Qy        522 CCACACCGTCCAACACATCACCACCCTCTGCCAACAACAAGGCATCAAAACCAAAACCCT 581
              |     |   ||  | |  | ||  |     ||  | |    |  ||    ||    |||
Db        415 CACACCNNCCCCCCCCCCCCCCCNCNCNNCCCCCCCCCCCCCCCCCCACCCCCCCCCCCC 356

Qy        582 CCCCACCAACCACGCCTTCCACTCCCCCCACACCAACCCCATCCTCAACCAACTCCACCA 641
              |||| ||  ||  |  |  || | ||| |  | || | |||  || |  ||  | || ||
Db        355 CCCCCCCCCCCCCCACCACCCCCCCCNCACCCCCCAACCCCCCCCCCCCCCCCCCCCCCC 296

Qy        642 GCACACCCAAACCCTCACCTACCACCCACCCCACACCCCCCTCATCACCGCCAACACCCC 701
              | ||| |   ||| | ||  || ||| || | ||||| || |  | ||| ||  || ||||
Db        295 ACCCACACNCCCCCCCCCCCCCCCCCCCCCCNCCCACCCACCCCCCCCCCCCCCCACCCCCC 236

Qy        702 ACCCGACCAACTCCTCACCCCCCACTACTGGACCCAACAAGCCCGCAACACCGTCGACTA 761
              | |  ||  | || ||||||||||  |    | ||||  ||| || || | | |
Db        235 CCNCCCCCCCCCCCNCACCCCCCACCCCCCCCCACAAC--CCCCCCACCAACCCCCCCC 178

Qy        762 CGCCACCACCACCCAAACCCTCCACCAACACGGCGTCACCACCTACATCGAACTCGGACC 821
              |  || |||| ||| ||| || || || |  |  |||| ||  || |   | |   ||
Db        177 ACCNCCCCCCACNCAACCCCCCCCCCCACCCCCCCCCACCCCCCCCACCCCCCCCCCCCC 118

Qy        822 CGACAACACCCTCACCACCCTCACCCACCACAACCTCCCCAACCCCCCCACCACCACCCT 881
              |  |  | ||| | || ||  | |||| | |  || ||||  ||| ||| || || |||
Db        117 CCCCCNCCCCCCCCCCCCCCACCCCCCANCCCCCCCCCCCCCCCCCCNCCCCCCCCCCCCC 58

Qy        882 CACCCTCACCCACCCCCACCACCACCCCCAAACCCACCTCCTC 924
              | ||| | ||| ||||| || || |||||   ||| || || |
Db         57 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 15
```

```
RESULT 7
AG152606/c
LOCUS       AG152606               1138 bp    DNA     linear   GSS 09-JAN-2002
DEFINITION  Pan troglodytes DNA, clone: RP43-016B03.TJ, genomic survey
            sequence.
ACCESSION   AG152606
VERSION     AG152606.1  GI:16682284
KEYWORDS    GSS.
SOURCE      Pan troglodytes male lymphocytes DNA, clone_lib:RPCI-43 Chimpanzee
            Male BAC Library clone:RP43-016B03.TJ.
  ORGANISM  Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE   1
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     BAC end sequences of Library RPCI-43
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1138)
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            (E-mail:chimpbes@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT     Clones are derived from the chimpanzee BAC library RPCI-43 This BAC
            end was generated during the R&D process and may have higher chance
            of clone tracking errors.
             PRIMERS
            Sequencing: TJ
            LIBRARY
            Vector    : pBACe3.6
            R.Site 1  : EcoRI
            R.Site 2  : EcoRI.
FEATURES            Location/Qualifiers
     source         1. .1138
                    /organism="Pan troglodytes"
                    /db_xref="taxon:9598"
                    /clone="RP43-016B03.TJ"
                    /sex="male"
                    /cell_type="lymphocytes"
                    /clone_lib="RPCI-43 Chimpanzee Male BAC Library"
BASE COUNT      38 a      9 c    869 g     38 t    184 others
ORIGIN
```

Query Match 25.6%; Score 237; DB 17; Length 1138;
Best Local Similarity 49.3%; Pred. No. 6.9e-29;
Matches 441; Conservative 0; Mismatches 454; Indels 0; Gaps 0;

```
Qy         27 CACCCAACGCCCCGGCATGGCCCACGGCCTCTACCACACCCACCCCGTCTTCGCCGCCGC 86
              | |||  | ||||  |    ||| |  || |  || |  || ||||  |  | || || |
Db       1106 CCCCCCCCCCCCCCCCNNCCCCCCCCCCCNCCNCCCCCNCCCCCCCCCCCCCCCCCCCC 1047

Qy         87 ACTCAACGACATCTGCACCCACCTCGACCCCCACCTCGACCACCCCCTCCTCCCCCTCCT 146
              | |  |  |  |  | ||| || |  |||||  | |  || ||||| || ||||| ||
Db       1046 CCCCCCCCNCCCCCCCCCCCCCCCCCCCCCCCCNNCCCCCCCCCCCCCCCCCCCCCCCCC 987

Qy        147 CACCCAAAACGACAACGACAACGAGGACGCGGCCGCACTGCTCCAGCAGACCCGCTACGC 206
              | |||    |  |  |  |  |    |  |  || | |  | || |   ||| |  | |
Db        986 CCCCCCCCCCCCCCCCCCCCCCCCNNCCCCNCCCCCCCCCCCCCCCCCCCCCCCCCCCC 927

Qy        207 CCAGCCCGCCCTCTTCGCCTTCCAGGTCGCCCTCCACCGCCTCCTCACCGACGGCTACCA 266
              ||  |||  || |  | ||  ||    | ||| || || ||    | ||  |  |  ||
Db        926 CCCNCCCCNCCCCCCCCCCCCCCCCNCCCCCCCNCCCCCCCCCNNCCCCCCCCCCCCCCC 867

Qy        267 CATCACCCCCCACTACTACGCCGGACACTCCCTCGGCGAAATCACCGCCGCCCACCTCGC 326
              |     |||||| |  |  | ||   | | ||| |  |      | || || ||| || | |
Db        866 CCCNCCCCCCCNCCCCCCCCCCCCCCCCCCCCCCCCCCCCNCCCCCCCCCCCCCCCCCCC 807

Qy        327 CGGCATCCTCACCCTCACCGACGCCACCACCCTCATCACCCAACGCGCCACCCTCATGCA 386
              |  |  || | ||| | ||  | ||| || ||| |     |||  | | || ||| |   |
Db        806 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCNNCCCCCCCCCCCCCCCCCCCC 747

Qy        387 AACCATGCCCCCCGGCACCATGACCACCCTCCACACCACCCCCCACCACATCACCCACCA 446
                ||   ||| |    ||       |||||| | || ||||||  | |  | ||| ||
Db        746 CCCCCCCCCCNCNCCNCCCNCCCNNCCCCTCCNCCCCCCCCCCCCCNCCCCCCCCCCCCC 687

Qy        447 CCTCACCGCCCACGAAAACGACCTCGCCATCGCCGCCATCAACACCCCCACCTCCCTCGT 506
              || |    |||  |      ||   ||  |  |      | | || ||| |  | ||| |
Db        686 CCCCNNNCCCCCCCNCNCCCCCCCNCCCCNCNNCNNNCCCCCCCCCCNCCCCCCCCCCCC 627

Qy        507 CATCAGCGGCACCCCCCACACCGTCCAACACATCACCACCCTCTGCCAACAACAAGGCAT 566
              |  |  |  | |||| | | ||  ||  | |  | || ||| |  ||  |  |    |
Db        626 CCCCCNCCCCCCCCCNCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 567

Qy        567 CAAAACCAAAACCCTCCCCACCAACCACGCCTTCCACTCCCCCCACACCAACCCCATCCT 626
              |    ||    ||| |||| ||  || | ||  |  | |      | |  || |||  ||
Db        566 CCCCCCCCCCCCCCCCCCCCNCCNCCCCCCCCCCNCNNCACAAAAAAAAAAAAAACCCCNCCC 507

Qy        627 CAACCAACTCCACCAGCACACCCAAACCCTCACCTACCACCCACCCCACACCCCCTCAT 686
              |  ||  | || ||    | |||    |||   ||  || || |  | || |  |||
Db        506 CCCCCTCCNCCCCCCCNCCCCCCNCCCCCCNCCCCNCCCCCCNCCNCCCCCCCCCCCNNN 447

Qy        687 CACCGCCAACACCCCACCCGACCAACTCCTCACCCCCCCACTACTGGACCCAACAAGCCCG 746
              | || ||  | |||| |||  ||  | || | |||||| |        | | |   |||
Db        446 CNCCNCCNNCCCCCCCCCCCCCCCCCCCCCCNCCCCCCCCCNNNCCCCNCCCCCCCCCCC 387

Qy        747 CAACACCGTCGACTACGCCACCACCACCCAAACCCTCCACCAACACGGCGTCACCACCTA 806
              | | ||  |  |  | || || || ||    ||| || ||  | |  |        ||
Db        386 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCNCCCCCCCCCCCCCCCCCCCCCNNNNNNCCCC 327

Qy        807 CATCGAACTCGGACCCGACAACACCCTCACCACCCTCACCCACCACAACCTCCCCAACCC 866
```

```
Db        326 CCCCCNNCCCCCNNNCCCCCCCCCCNCCCCCNCCCNNNCNNNCCNNNCNCCCNNCCNCCN 267

Qy        867 CCCCACCACCACCCTCACCCTCACCCACCCCCCACCACCACCCCCCAAACCCACCTC 921

Db        266 NCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCNNNCCCNCCNNCNNNNNNNNNC 212


RESULT 8
BE455154/c
LOCUS       BE455154               1082 bp    mRNA    linear   EST 22-OCT-2001
DEFINITION  HVSMEh0096G08f Hordeum vulgare 5-45 DAP spike EST library
            HVcDNA0009 (5 to 45 DAP) Hordeum vulgare cDNA clone HVSMEh0096G08f,
            mRNA sequence.
ACCESSION   BE455154
VERSION     BE455154.2  GI:16322259
KEYWORDS    EST.
SOURCE      Hordeum vulgare.
  ORGANISM  Hordeum vulgare
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
            ; Triticeae; Hordeum.
REFERENCE   1  (bases 1 to 1082)
  AUTHORS   Wing,R., Close,T.J., Kleinhofs,A., Wise,R., Begum,D., Frisch,D., Yu
            ,Y., Henry,D., Palmer,M., Rambo,T., Simmons,J., Choi,D.W., Fenton
            ,R.D., Close,S.J., Oates,R. and Main,D.
  TITLE     Development of a genetically and physically anchored EST resource
            for barley genomics: Morex 5-45 DAP spike cDNA library
  JOURNAL   Unpublished (2001)
COMMENT     On Jul 26, 2000 this sequence version replaced gi:9464545.
            Contact: Wing RA
            Clemson University Genomics Institute
            Clemson University
            100 Jordan Hall, Clemson, SC 29634, USA
            Tel: 864 656 7288
            Fax: 864 656 4293
            Email: rwing@clemson.edu
            Total hq bases = 280
            Seq primer: AATTAACCCTCACTAAAGGG
            High quality sequence start: 3
            High quality sequence stop: 1073.
FEATURES             Location/Qualifiers
     source          1. .1082
                     /organism="Hordeum vulgare"
                     /cultivar="Morex"
                     /db_xref="taxon:4513"
                     /clone="HVSMEh0096G08f"
                     /clone_lib="Hordeum vulgare 5-45 DAP spike EST library
                     HVcDNA0009 (5 to 45 DAP)"
                     /tissue_type="5-45 DAP Spike"
                     /lab_host="SOLR"
                     /note="Vector: lambdaZAP; Site_1: EcoR1; Site_2: Xho1;
                     Plants were grown in the greenhouse at the University of
                     California, Riverside (Fenton, SJ Close, TJ Close). Whole
                     spikes with awns trimmed were collected at 5, 10, 15, 20,
                     30 and 45 DAP (Fenton). Total RNA was prepared from each
                     pool, equal quantities of all six RNA pools were combined,
                     poly(A) RNA was purified from the mixture, one primary
                     unamplified cDNA library was made, and 1 million pfu were
                     in vivo excised to give pBluescript SK(-) cDNA phagemids
                     (Choi)in the TJ Close lab at the University of California,
                     Riverside. Phagemids were plated and picked at the Clemson
                     University Genomics Institute (CUGI) (Begum, Palmer,
                     Frisch, Atkins and Wing). Plasmid DNA preparations, DNA
                     sequencing and sequence analysis were performed at CUGI
                     (Wing, Yu, Frisch, Henry, Simmons, Oates, Rambo, Main).
                     The sequence has been trimmed to remove vector sequence
                     and contains a minimum of 100 bases of phred value 20 or
                     above. For more details on library preparation and
                     sequence analysis see
                     http://www.genome.clemson.edu/projects/barley. To order
                     this clone see http://www.genome.clemson.edu/orders Also
                     see Close TJ, Wing R, Kleinhofs A, Wise R (2001)
                     Genetically and physically anchored EST resources for
                     barley genomics. Barley Genetics Newsletter 31:29-30.
                     (http://wheat.pw.usda.gov/ggpages/bgn/31/cover.html)"
BASE COUNT       89 a     15 c    937 g     33 t      8 others
ORIGIN

  Query Match           25.6%;  Score 236.4;  DB 10;  Length 1082;
  Best Local Similarity  53.8%;  Pred. No. 8.7e-29;
  Matches  483;  Conservative    0;  Mismatches  415;  Indels    0;  Gaps    0;

Qy         27 CACCCAACGCCCCGGCATGGCCCACGGCCTCTACCACACCCACCCCGTCTTCGCCGCCGC 86

Db       1045 CCCCCCCCCCTTCCGCTCCATCCCCTCTCCTCCCCCCTCTCCCCCCACCCTCCCACCCCC 986

Qy         87 ACTCAACGACATCTGCACCCACCTCGACCCCCACCTCGACCACCCCCTCCTCCCCCTCCT 146

Db        985 NACCCCCCCCCCCCCCCCCCTCCCTCGCCTCCCCGCCCCCCCTCCCCCCCCCCCCCCCCCC 926

Qy        147 CACCCAAAACGACAACGACAACGAGGACGCGGCCGCACTGCTCCAGCAGACCCGCTACGC 206

Db        925 CCCTCCCGTCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCTCCCCCCTCCCCCCTCCCCCC 866

Qy        207 CCAGCCCGCCCTCTTCGCCTTCCAGGTCGCCCTCCACCGCCTCCTCACCGACGGCTACCA 266

Db        865 CCTCCTCCCCCTCTCCCCCCCCCGTCCCCCCCTCCCCCCCCTCCCCCCCCCCCTCGCCCC 806

Qy        267 CATCACCCCCCACTACTACGCCGGACACTCCCTCGGCGAAATCACCGCCGCCCACCTCGC 326

Db        805 CCCCCCCCCCCCCCCCCTCTCCCCCCTCTTCCCCCTTCCCCCCCCCCCCCCCCCCCCCCC 746

Qy        327 CGGCATCCTCACCCTCACCGACGCCACCACCCTCATCACCCAACGCGCCACCCTCATGCA 386

Db        745 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCTCCCCCCCCCCCCCCCCCTCCCCCTCC 686

Qy        387 AACCATGCCCCCCGGCACCATGACCACCCTCCACACCACCCCCCACCACATCACCCACCA 446

Db        685 CCCCCCCCCCTTCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCTCCTCCTCCCC 626

Qy        447 CCTCACCGCCCACGAAAACGACCTCGCCATCGCCGCCATCAACACCCCCACCTCCCTCGT 506

Db        625 CCCCCCTCTCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 566

Qy        507 CATCAGCGGCACCCCCCACACCGTCCAACACATCACCACCCTCTGCCAACAACAAGGCAT 566

Db        565 CCCCCCCCCCCCCCCCCCCCCTCCCCCCCCCACCCCCCCCCCCCCCCCCCCCCCCCCCC 506

Qy        567 CAAAACCAAAACCCTCCCCACCAACCACGCCTTCCACTCCCCCCACACCAACCCCATCCT 626

Db        505 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCTCTTCCCCCCCC 446

Qy        627 CAACCAACTCCACCAGCACACCCAAACCCTCACCTACCACCCACCCCACACCCCCCTCAT 686

Db        445 CCCCCCCCCCCCTCCCCCCCCCCTCCCCCCCCCCCCCCCCCCCCCCCCCCTCCCCCCCCC 386

Qy        687 CACCGCCAACACCCCACCCGACCAACTCCTCACCCCCCCACTACTGGACCCAACAAGCCCG 746

Db        385 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCAAAAAAACCCC 326

Qy        747 CAACACCGTCGACTACGCCACCACCACCCAAACCCTCCACCAACACGGCGTCACCACCTA 806

Db        325 CCCCCCCCCCCCCCCCCCCCCCCTCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 266

Qy        807 CATCGAACTCGGACCCGACAACACCCTCACCACCCTCACCCACCACAACCTCCCCAACCC 866

Db        265 CCCTCCCCCCCCCCCCCCCCTCCCCCCCCCCCCCCCCCCACACCCACCCACTCCCCCCCCN 206

Qy        867 CCCCACCACCACCCTCACCCTCACCCACCCCCACCACCACCCCCAAACCCACCTCCTC 924

Db        205 ACATCCCCACCCCCCCCCCCCCCCCCCCCCCCCCCCCCNCCCCCCCNCCCCCCCCCGC 148


RESULT 9
BI416537/c
```

```
LOCUS       BI416537               1224 bp    mRNA    linear   EST 15-AUG-2001
DEFINITION  hasp001xe15f Heterobasidion annosum - Scots pine infection stage
            subtraction cDNA library (hasp) Pinus sylvestris/Heterobasidion
            annosum cDNA clone hasp001xe15f, mRNA sequence.
ACCESSION   BI416537
VERSION     BI416537.1  GI:15187560
KEYWORDS    EST.
SOURCE      Pinus sylvestris/Heterobasidion annosum.
  ORGANISM  Pinus sylvestris/Heterobasidion annosum
            Eukaryota; mixed EST libraries.
REFERENCE   1  (bases 1 to 1224)
  AUTHORS   Asiegbu,F.O., Nahalkova,J., Choi,W., Stenlid,J. and Dean,R.A.
  TITLE     Expressed sequence tags of randomly selected cDNA clones from the
            interaction of the root rot fungus (Heterobasidion annosum) with
            seedling roots of Scots pine (Pinus sylvestris)
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Fred O. Asiegbu
            Dept. of Forest Mycology & Pathology
            Swedish University of Agriculture, Box 7026,S-750 07, Uppsala,
            Sweden
            Tel: +46 18 67 15 98
            Fax: +46 18 30 92 45
            Email: Fred.Asiegbu@mykopat.slu.se
            Seq primer: T7 primer.
FEATURES             Location/Qualifiers
     source          1. .1224
                     /organism="Pinus sylvestris/Heterobasidion annosum"
                     /db_xref="taxon:169015"
                     /clone="hasp001xe15f"
                     /clone_lib="Heterobasidion annosum - Scots pine infection
                     stage subtraction cDNA library (hasp)"
                     /dev_stage="Seedling roots of scots pine were infected for
                     6 days with H. annosum"
                     /note="Vector: pT-Adv; Site_1: EcoRI; The subtractive
                     hybridization cDNA library was constructed from  scots
                     pine roots infected for 6-days with mycelia of
                     Heterobasidon  annosum (FP5)."
BASE COUNT       50 a      14 c    1063 g      38 t      59 others
ORIGIN

  Query Match           25.6%;  Score 236.2;  DB 13;  Length 1224;
  Best Local Similarity   53.2%;  Pred. No. 9.3e-29;
  Matches  479;  Conservative    0;  Mismatches  419;  Indels    3;  Gaps    1;

Qy         27 CACCCAACGCCCCGGCATGGCCCACGGCCTCTACCACACCCACCCCGTCTTCGCCGCCGC 86
              | |||  | ||||  |    ||| |  || |  || ||||||||||  |  | || || |
Db       1144 CCCCCCCCACCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCACCCACCCCCCCCCCCCCCC 1085

Qy         87 ACTCAACGACATCTGCACCCACCTCGACCCCCACCTCGACCACCCCCTCCTCCCCCTCCT 146
                |  ||  |  | ||| || |  | ||| || |  || |||||  || ||||| ||
Db       1084 CCCTCCCCACCCCCCCCCCCCCCCCCCCCTCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 1025

Qy        147 CACCCAAAACGACAACGACAACGAGGACGCGGCCGCACTGCTCCAGCAGACCCGCTACGC 206
              |||||    |  | | |  | |     | |  || | |  | ||  |   ||| |  | |
Db       1024 CACCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 965

Qy        207 CCAGCCCGCCCTCTTCGCCTTCCAGGTCGCCCTCCACCGCCTCCTCACCGACGGCTACCA 266
              ||  ||| ||| |  | ||  ||    | ||| || || || || | ||  |  |  ||
Db        964 CCTCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 905

Qy        267 CATCACCCCCCACTACTACGCCGGACACTCCCTCGGCGAAATCACCGCCGCCCACCTCGC 326
              |  | ||||||   ||  | ||   | | ||| |  |      | || || ||| ||    |
Db        904 CCCCCCCCCCCCACACCCCCCCCCTCTCCCCCCCCCCCCCCCCCCCCCCCCCCCNTCC 845

Qy        327 CGGCATCCTCACCCTCACCGACGCCACCACCCTCATCACCCAACGCGCCACCCTCATGCA 386
              |  || || | |||  | ||  | ||||  ||| |  | | |  | | ||  ||  |   |
Db        844 CCCCACCCCCCCCCCCCCCCCCCCCCACNCCCCCCCCCCCACCCCCCCCCCCCNNCCCACC 785

Qy        387 AACCATGCCCCCC---GGCACCATGACCACCCTCCACACCACCCCCCCACCACATCACCCA 443
                || | ||||||     | || |  || ||| || |||| |||||| || |  | |||
Db        784 CCCCCTCCCCCCCCANCCCCCCCCTCTCCCCCCCCCCCACCCCCCCCCCCCCCCCCCCCC 725

Qy        444 CCACCTCACCGCCCACGAAAACGACCTCGCCATCGCCGCCATCAACACCCCCACCTCCCT 503
```

```
                || | || || ||| |     |  || |  |  | || |.  || | ||||| || |||
Db        724 CCCCNTCCCCCCCCCCCCCCCCCCCCCCCCCACCCCCCCCCNTCCACCCCCCCCCCCCCCC 665

Qy        504 CGTCATCAGCGGCACCCCCCACACCGTCCAACACATCACCACCCTCTGCCAACAACAAGG 563
              |   | |  |  | | |||| | ||  |    |    | | ||| |  || ||  |
Db        664 CCCNACCCCCCCCCCNCCCCCCCCCCNCTCCCCACCCCCNTCCCCCACCCCACNCCTCTC 605

Qy        564 CATCAAAACCAAAACCCTCCCCACCAACCACGCCTTCCACTCCCCCCACACCAACCCCAT 623
              || |    ||    ||  |||| ||   | | | |||| |  |||   | ||  ||||
Db        604 CACCCCCCCCTCCCCCANCCCCCCCCCNCCCCCNTTCCCCCNCCCNTCCCCCCCCCCCCC 545

Qy        624 CCTCAACCAACTCCACCAGCACACCCAAACCCTCACCTACCACCCACCCCACACCCCCCT 683
              |    |||  | || ||  | | |||   ||||| ||  || ||| | || | ||||||
Db        544 CTCANACCTCCCCCCCCCACCCCCCCTNCCCCTCCCCCCCCCCCCCCCTCCCCCCCCCCC 485

Qy        684 CATCACCGCCAACACCCCACCCGACCAACTCCTCACCCCCCACTACTGGACCCAACAAGC 743
                |||  ||  | | || |||  ||    || | ||||||  |  |    |||  |
Db        484 ACCCACNTCCCCCCCTCCCCCCCCCCNNTCCCCCCCCCCCCCCCCCCCCCCCCCCCCCT 425

Qy        744 CCGCAACACCGTCGACTACGCCACCACCACCCAAACCCTCCACCAACACGGCGTCACCAC 803
              || |  | | ||  |    |  |  || |||   |||||| ||    |  |  | ||
Db        424 CCCCTCCCTCCTCCCCNCNCCNCCNCCCTCCCCCCCCCTCCCCCCCTNCCCCCNCCCCCN 365

Qy        804 CTACATCGAACTCGGACCCGACAACACCCTCACCACCCTCACCCACCACAACCTCCCCAA 863
              |  |  |    |   |||  |  | ||| | || ||  | |||  | |  |||||||
Db        364 CCCCCCCCNCNCCCCCCCCCCCCCCCCCCCCCCCCCCCCCNCCCCCCNNCCCCCCCTCCCCCC 305

Qy        864 CCCCCCCACCACCACCCTCACCCTCACCCACCCCCACCACCACCCCCAAACCCACCTCCT 923
              ||||| | || || | | | ||| | |  | | || || ||||    |  || ||
Db        304 CCCCCTCCCCCCCCCNCCCCCCCCCCCCGGCCTCTCCCCNCCCCCCCCCCCACACCCNCCC 245

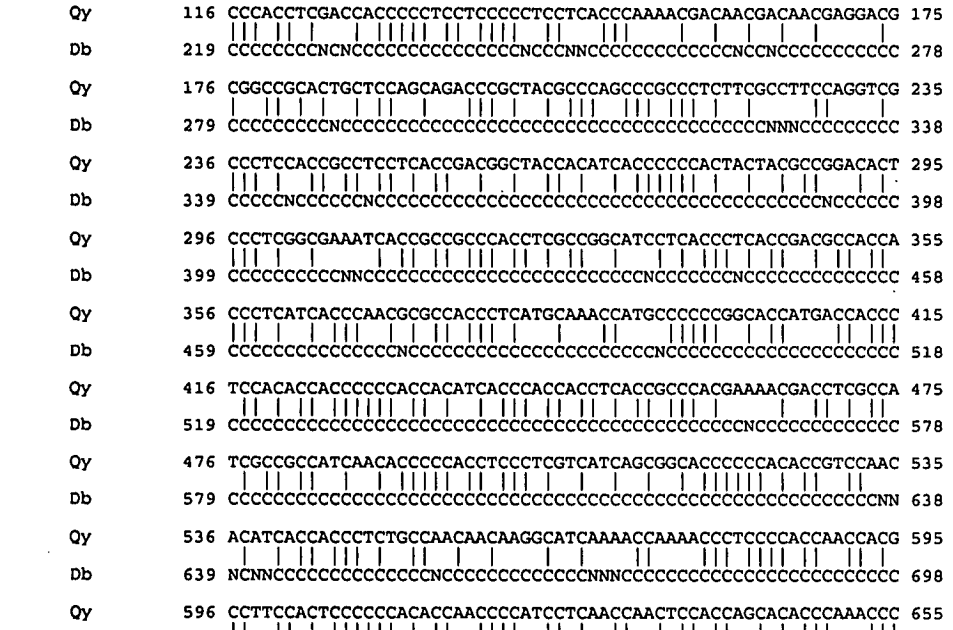
Qy        924 C 924
              |
Db        244 C 244
```

```
RESULT 10
BM912505
LOCUS       BM912505               1161 bp    mRNA    linear   EST 12-MAR-2002
DEFINITION  AGENCOURT_6612272 NIH_MGC_41 Homo sapiens cDNA clone IMAGE:5474051
            5', mRNA sequence.
ACCESSION   BM912505
VERSION     BM912505.1  GI:19362884
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1161)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: DCTD/DTP
             cDNA Library Preparation: Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1986  row: k  column: 12
            High quality sequence stop: 218.
FEATURES             Location/Qualifiers
     source          1. .1161
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5474051"
                     /clone_lib="NIH_MGC_41"
                     /tissue_type="amelanotic melanoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: skin; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming. Directionally cloned
```

into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH_MGC Library."

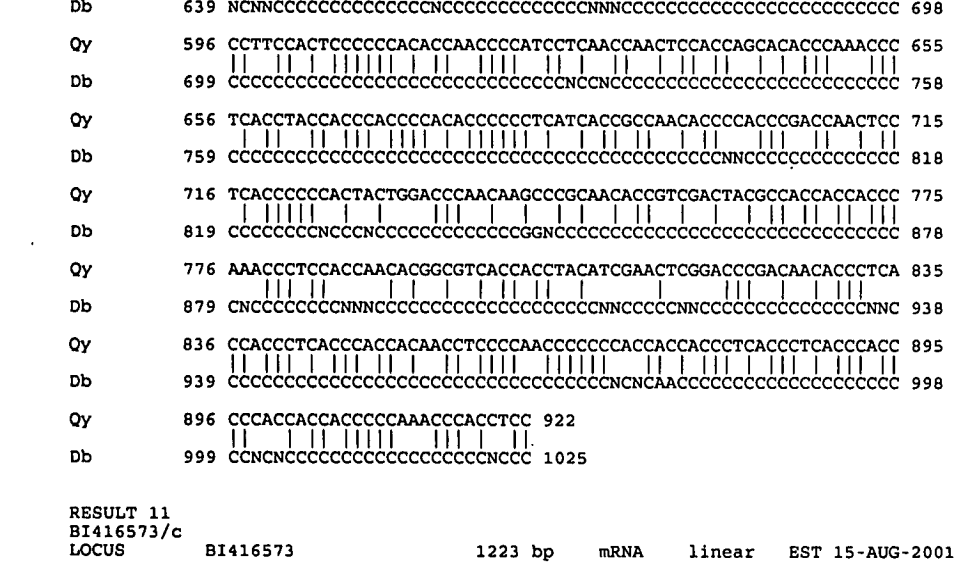
BASE COUNT 53 a 844 c 107 g 88 t 69 others
ORIGIN

Query Match 25.4%; Score 235; DB 14; Length 1161;
Best Local Similarity 53.3%; Pred. No. 1.4e-28;
Matches 430; Conservative 0; Mismatches 377; Indels 0; Gaps 0;



RESULT 11
BI416573/c
LOCUS BI416573 1223 bp mRNA linear EST 15-AUG-2001
DEFINITION hasp001xg11f Heterobasidion annosum - Scots pine infection stage subtraction cDNA library (hasp) Pinus sylvestris/Heterobasidion annosum cDNA clone hasp001xg11f, mRNA sequence.
ACCESSION BI416573
VERSION BI416573.1 GI:15187596
KEYWORDS EST.
SOURCE Pinus sylvestris/Heterobasidion annosum.
ORGANISM Pinus sylvestris/Heterobasidion annosum
Eukaryota; mixed EST libraries.
REFERENCE 1 (bases 1 to 1223)
AUTHORS Asiegbu,F.O., Nahalkova,J., Choi,W., Stenlid,J. and Dean,R.A.
TITLE Expressed sequence tags of randomly selected cDNA clones from the interaction of the root rot fungus (Heterobasidion annosum) with seedling roots of Scots pine (Pinus sylvestris)
JOURNAL Unpublished (2001)
COMMENT Contact: Fred O. Asiegbu
Dept. of Forest Mycology & Pathology
Swedish University of Agriculture, Box 7026,S-750 07, Uppsala, Sweden
Tel: +46 18 67 15 98
Fax: +46 18 30 92 45
Email: Fred.Asiegbu@mykopat.slu.se
Seq primer: T7 primer.
FEATURES Location/Qualifiers
source 1. .1223
/organism="Pinus sylvestris/Heterobasidion annosum"
/db_xref="taxon:169015"
/clone="hasp001xg11f"
/clone_lib="Heterobasidion annosum - Scots pine infection stage subtraction cDNA library (hasp)"
/dev_stage="Seedling roots of scots pine were infected for 6 days with H. annosum"
/note="Vector: pT-Adv; Site_1: EcoRI; The subtractive hybridization cDNA library was constructed from scots pine roots infected for 6-days with mycelia of Heterobasidon annosum (FP5)."
BASE COUNT 49 a 24 c 1047 g 34 t 69 others
ORIGIN

Query Match 25.2%; Score 232.8; DB 13; Length 1223;
Best Local Similarity 55.5%; Pred. No. 3.2e-28;
Matches 457; Conservative 0; Mismatches 364; Indels 2; Gaps 2;



```
Db        798 CCCCCCCCAACCCCCCCCCCNCCCCCCCCCCCCCCCCCCCCCCCCCCTCCCCCCCCCCCC 739

Qy        581 TCCCCACCAACCACGCCTTCCACTCCCCCCACACCAACCCCATCCTCAACCAACTCCACC 640
              ||||| ||  || | |   ||   ||| ||  |||  ||||| || |  ||  | || ||
Db        738 TCCCCCCCCCCCCCCCTCCCCNTCCCCTCCCNACCCCCCCCACCCNCCCCCCNCCCCCCC 679

Qy        641 AGCACACCCAAACCCTCACCTACCACCCACCCCACACCCCCCTCATCACCGCCAACACCC 700
                | |||||   ||| | ||  || ||| |||| ||| |||  | || || ||     |||
Db        678 TCCCCACCCCCCCCCCCCCCNCCCCCCCCCCCCCCCACNCCCACCCTCTCCCCCCCATCCC 619

Qy        701 CACCCGACCAACTCCTCACCCCCCACTACTGGACCCAACAAGCCCGCAACACCGTCGACT 760
              |||||  ||  | || | | |||| |  |    ||  |  | |  | |  |  |  ||
Db        618 CACCC-CCCNNCCCCNCTCTCCCCCCCCCACCCCCTCNCCCCCTCCNACCCACNCCCCCT 560

Qy        761 ACGCCACCACCACCCAAACCCTCCACCAACACGGCGTCACCACCTACATCGAACTCGGAC 820
              | ||  |  | |||| ||||||| ||  | |  |  |  | ||| |  |   | |   |
Db        559 CCTCCCACCACCCCCACACCCTCCNCCCCCCCCCCCCCCCNCCCCTCCCCCCCCCCCCCCC 500

Qy        821 CCGACAACACCCTCACCACCCTCACCCACCACAACCTCCCCAACCCCCCCACCACCACCC 880
              ||    | ||||| || ||| | ||| || |   || ||||  ||||||| || || |||
Db        499 CCCCTCCCCCCCTCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 440

Qy        881 TCACCCTCACCCACCCCCACCACCACCCCCAAACCCACCTCCT 923
                |  || | ||| ||||| || |  |||||    ||| || |||
Db        439 CCCNCCCCCCCCCCCCCCCCCCCCCTCCCCCCCCCCCCCCCCCT 397


RESULT 12
BF254665/c
LOCUS       BF254665               1036 bp    mRNA    linear   EST 22-OCT-2001
DEFINITION  HVSMEf0004K19f Hordeum vulgare seedling root EST library HVcDNA0007
            (Etiolated and unstressed) Hordeum vulgare cDNA clone
            HVSMEf0004K19f, mRNA sequence.
ACCESSION   BF254665
VERSION     BF254665.2  GI:16315121
KEYWORDS    EST.
SOURCE      Hordeum vulgare.
  ORGANISM  Hordeum vulgare
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
            ; Triticeae; Hordeum.
REFERENCE   1  (bases 1 to 1036)
  AUTHORS   Wing,R., Close,T.J., Kleinhofs,A., Wise,R., Begum,D., Frisch,D., Yu
            ,Y., Henry,D., Palmer,M., Rambo,T., Simmons,J., Choi,D.W., Fenton
            ,R.D., Oates,R. and Main,D.
  TITLE     Development of a genetically and physically anchored EST resource
            for barley genomics: Morex unstressed seedling root cDNA library
  JOURNAL   Unpublished (2001)
COMMENT     On Nov 16, 2000 this sequence version replaced gi:11183770.
            Contact: Wing RA
            Clemson University Genomics Institute
            Clemson University
            100 Jordan Hall, Clemson, SC 29634, USA
            Tel: 864 656 7288
            Fax: 864 656 4293
            Email: rwing@clemson.edu
            Total hq bases = 200
            Seq primer: AATTAACCCTCACTAAAGGG
            High quality sequence start: 24
            High quality sequence stop: 1004.
FEATURES             Location/Qualifiers
     source          1. .1036
                     /organism="Hordeum vulgare"
                     /cultivar="Morex"
                     /db_xref="taxon:4513"
                     /clone="HVSMEf0004K19f"
                     /clone_lib="Hordeum vulgare seedling root EST library
                     HVcDNA0007 (Etiolated and unstressed)"
                     /tissue_type="Seedling root"
                     /lab_host="TJC121"
                     /note="Vector: lambdaZAP; Site_1: EcoR1; Site_2: Xho1;
                     Seeds were surface sterilized then germinated under axenic
                     conditions in the dark at room temperature on filter paper
                     with water, nystatin and cefotaxime in covered
                     crystallization dishes. Five-day old seedling roots were
                     then harvested, total RNA was prepared, poly(A) RNA was
                     purified, one primary unamplified cDNA library was made,
                     and 1 million pfu were in vivo excised to give pBluescript
                     SK(-) cDNA phagemids. These steps were performed in the TJ
                     Close laboratory at the University of California,
                     Riverside (Choi, Close, Fenton). Phagemids were plated and
                     picked at the Clemson University Genomics Institute (CUGI)
                     (Begum, Palmer, Frisch, Atkins and Wing). Plasmid DNA
                     preparations, DNA sequencing and sequence analysis were
                     performed at CUGI (Wing, Yu, Frisch, Henry, Simmons, Oates
                     , Rambo, Main). The sequence has been trimmed to remove
                     vector sequence and contains a minimum of 100 bases of
                     phred value 20 or above. For more details on library
                     preparation and sequence analysis see
                     http://www.genome.clemson.edu/projects/barley. To order
                     this clone see http://www.genome.clemson.edu/orders Also
                     see Close TJ, Wing R, Kleinhofs A, Wise R (2001)
                     Genetically and physically anchored EST resources for
                     barley genomics. Barley Genetics Newsletter 31:29-30.
                     (http://wheat.pw.usda.gov/ggpages/bgn/31/cover.html)"
BASE COUNT       68 a     43 c    858 g     50 t     17 others
ORIGIN

  Query Match             24.9%;  Score 230;  DB 12;  Length 1036;
  Best Local Similarity   53.7%;  Pred. No. 9.2e-28;
  Matches  473;  Conservative    0;  Mismatches  408;  Indels    0;  Gaps    0;

Qy         42 CATGGCCCACGGCCTCTACCACACCCACCCCGTCTTCGCCGCCGCACTCAACGACATCTG 101
              | |  |||    || |  || |  |  ||||  |  | || || ||| |  |  | |
Db       1016 CCTCTCCCCACCCCCCCCCCCCTCACCCCCCCCCCCTCCCCCCCACCCCTCCCCCCCCC 957

Qy        102 CACCCACCTCGACCCCCACCTCGACCACCCCCTCCTCCCCCTCCTCACCCAAAACGACAA 161
              | ||  || |  ||||| ||    || ||||| |||||||| || | |||    |  |
Db        956 CCCCGCCCCCTCCCCCCCCCCGCCCCCCCCCCCACCTCCCCCCCCCCCCCCCCCTCCCCCCC 897

Qy        162 CGACAACGAGGACGCGGCCGCACTGCTCCAGCAGACCCGCTACGCCCAGCCCGCCCTCTT 221
              | ||  |    | |  | | |  ||||      ||| || | |||| ||| | | |
Db        896 CCACTCCTCTCCCCCCCTCCCCCTCTCCCCTCCCCCCCCTCCCCCCACCCCCCACGCCC 837

Qy        222 CGCCTTCCAGGTCGCCCTCCACCGCCTCCTCACCGACGGCTACCACATCACCCCCCACTA 281
              | ||  ||    | | | || || || || | ||  |  |  || |  | ||  || |
Db        836 CCCCCCCCCCCCCCCTCCCCCCCCCCCCCCCCCCCCCGCCCCCCCCCCCCCCCCCTACCCCCC 777

Qy        282 CTACGCCGGACACTCCCTCGGCGAAATCACCGCCGCCCACCTCGCCGGCATCCTCACCCT 341
              | ||   | | ||| |  |      | || || ||| || | ||  |  || | |||
Db        776 GACCCCCCCCCTCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 717

Qy        342 CACCGACGCCACCACCCTCATCACCCAACGCGCCACCCTCATGCAAACCATGCCCCCCGG 401
              | ||  | ||  | ||| |  | |||  | | || ||| |  |   ||   || ||
Db        716 CCCCCCCCCCCTCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCACCACC 657

Qy        402 CACCATGACCACCCTCCACACCACCCCCCACCACATCACCCACCACCTCACCGCCCACGA 461
              | ||    || ||| || | || | |  | || |  | ||  || || | || ||| |
Db        656 CCCCCACCCCCCCCCCCCCTCCGCTCAGCCCCCCCCCCCCCGTCCCCCCCCCCCCCCCCCC 597

Qy        462 AAACGACCTCGCCATCGCCGCCATCAACACCCCCACCTCCCTCGTCATCAGCGGCACCCC 521
                |  || | ||    || ||  | | ||||| || ||  |  | |  |    ||||
Db        596 CCCCCGCCGCCCCCCGCCCCCCCACCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCTCCCCC 537

Qy        522 CCACACCGTCCAACACATCACCACCCTCTGCCAACAACAAGGCATCAAAACCAAAACCCT 581
              || | ||  ||  | || | || ||| |  ||| |  |   |  |   ||   |||
Db        536 CCCCCCCCCCCCCCCCACCCCCCCCCCCCCCCCACCCCCCCCCCCCCCCCCCCCCCCCCC 477

Qy        582 CCCCACCAACCACGCCTTCCACTCCCCCCACACCAACCCCATCCTCAACCAACTCCACCA 641
              || | ||  || | ||  || | |||||| ||||  ||||  |  |  ||  |    |
Db        476 CCNCCCCCCCCCCCCCCCCCCCCCCCCCCCCCACCCCCCCCCCCCGCCCCCCCCCCAGCCAC 417

Qy        642 GCACACCCAAACCCTCACCTACCACCCACCCCACACCCCCCTCATCACCGCCAACACCCC 701
                || ||| ||||| | ||  || ||| |||     ||||||  |  | || ||  | ||||
Db        416 CAACCCCCCAACCCCCCCCCCCCCCCCCCCCCGCGCCCCCCCCCCCCCCCCCCCCCCCCCC 357
```

```
Qy        702 ACCCGACCAACTCCTCACCCCCCCACTACTGGACCCAACAAGCCCGCAACACCGTCGACTA 761
              |||  || ||  | | ||||||    |    |||  |   |   ||   ||  | |
Db        356 CCCCCCCCCACCACCCCCCCCCCCACCCCCCCCCCCCCCCCACGAACACACCCCCCCCCCC 297

Qy        762 CGCCACCACCACCCAAACCCTCCACCAACACGGCGTCACCACCTACATCGAACTCGGACC 821
              | || || || |||   ||| || ||  ||    |  | ||   | | ||      | | ||
Db        296 CCCCCCCCCCCCCCCCCCCCCCCCGCCCCCAAAACACCCCCTTTTGCTTCCCGGGCCGNCC 237

Qy        822 CGACAACACCCTCACCACCCTCACCCACCACAACCTCCCCAACCCCCCCACCACCACCCT 881
              |   |  | ||| | || ||| | ||| || |  || ||||  ||||||| || || |||
Db        236 CNGCGCCCCCCGCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 177

Qy        882 CACCCTCACCCACCCCCACCACCACCCCCAAACCCACCTCC 922
              | ||| | ||| ||||| || || |||||   ||| || ||
Db        176 CCCCCGCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 136
```

RESULT 13
BF256755/c
LOCUS BF256755 1196 bp mRNA linear EST 22-OCT-2001
DEFINITION HVSMEf0010017f Hordeum vulgare seedling root EST library HVcDNA0007 (Etiolated and unstressed) Hordeum vulgare cDNA clone HVSMEf0010017f, mRNA sequence.
ACCESSION BF256755
VERSION BF256755.2 GI:16315528
KEYWORDS EST.
SOURCE Hordeum vulgare.
ORGANISM Hordeum vulgare
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae ; Triticeae; Hordeum.
REFERENCE 1 (bases 1 to 1196)
AUTHORS Wing,R., Close,T.J., Kleinhofs,A., Wise,R., Begum,D., Frisch,D., Yu ,Y., Henry,D., Palmer,M., Rambo,T., Simmons,J., Choi,D.W., Fenton ,R.D., Oates,R. and Main,D.
TITLE Development of a genetically and physically anchored EST resource for barley genomics: Morex unstressed seedling root cDNA library
JOURNAL Unpublished (2001)
COMMENT On Nov 16, 2000 this sequence version replaced gi:11185868.
Contact: Wing RA
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 7288
Fax: 864 656 4293
Email: rwing@clemson.edu
Total hq bases = 182
Seq primer: AATTAACCCTCACTAAAGGG
High quality sequence start: 95
High quality sequence stop: 1129.
FEATURES Location/Qualifiers
source 1. .1196
/organism="Hordeum vulgare"
/cultivar="Morex"
/db_xref="taxon:4513"
/clone="HVSMEf0010017f"
/clone_lib="Hordeum vulgare seedling root EST library HVcDNA0007 (Etiolated and unstressed)"
/tissue_type="Seedling root"
/lab_host="TJC121"
/note="Vector: lambdaZAP; Site_1: EcoR1; Site_2: Xho1; Seeds were surface sterilized then germinated under axenic conditions in the dark at room temperature on filter paper with water, nystatin and cefotaxime in covered crystallization dishes. Five-day old seedling roots were then harvested, total RNA was prepared, poly(A) RNA was purified, one primary unamplified cDNA library was made, and 1 million pfu were in vivo excised to give pBluescript SK(-) cDNA phagemids. These steps were performed in the TJ Close laboratory at the University of California, Riverside (Choi, Close, Fenton). Phagemids were plated and picked at the Clemson University Genomics Institute (CUGI) (Begum, Palmer, Frisch, Atkins and Wing). Plasmid DNA preparations, DNA sequencing and sequence analysis were performed at CUGI (Wing, Yu, Frisch, Henry, Simmons, Oates , Rambo, Main). The sequence has been trimmed to remove vector sequence and contains a minimum of 100 bases of phred value 20 or above. For more details on library preparation and sequence analysis see http://www.genome.clemson.edu/projects/barley. To order this clone see http://www.genome.clemson.edu/orders Also see Close TJ, Wing R, Kleinhofs A, Wise R (2001) Genetically and physically anchored EST resources for barley genomics. Barley Genetics Newsletter 31:29-30. (http://wheat.pw.usda.gov/ggpages/bgn/31/cover.html)"
BASE COUNT 75 a 35 c 948 g 105 t 33 others
ORIGIN

Query Match 24.8%; Score 229.2; DB 12; Length 1196;
Best Local Similarity 54.3%; Pred. No. 1.2e-27;
Matches 485; Conservative 0; Mismatches 392; Indels 16; Gaps 1;

```
Qy         27 CACCCAACGCCCCGGCATGGCCCACGGCCTCTACCACACCCACCCCGTCTTCGCCGCCGC 86
              | |||  |  |||  |   |||| || ||  |  || | ||||| ||  ||  |.|  ||||
Db       1131 CCCCCCCCCACCCCCCCCCGCCCCCGCCCACCCCCTCCCCCACACCCCCTCCTCGCCCGC 1072

Qy         87 ACTCAACGACATCTGCACCCACCTCGACCCCCACCTCGACCACCCCCTCCTCCCCCTCCT 146
              | |  |  |  || | ||| ||     ||||| |  |  || | ||| || ||||||||
Db       1071 CCCCCCCCCCCACTCCCCCCCCCCTCCCCCCCTCACCCCCCCCCACCCCCCCCCCCTCCC 1012

Qy        147 CACCCAAAACGACAACGACAACGAGGACGCGGCCGCACTGCTCCAGCAGACCCGCTACGC 206
              || |   |  |  |  |        || || |    | || || |    ||| |  |||
Db       1011 TCCCAACCGCCCCCTCCCCTCTCCAAACCCGCCACACCCGCCCCCGGCCTCCCCCCTCGC 952

Qy        207 CCAGCCCGCCCTCTTCGCCTTCCAGGTCGCCCTCCACCGCCTCCTCACCGACGGCTACCA 266
              |  ||| |||    | || ||||  || ||  || ||  | || | ||| |  |
Db        951 TCCCCCCCCCCCTCCCCCCCTCCACCTCCCCACCCCCCCACACCCCCCCGCCCCCGTTC 892

Qy        267 CATCACCCCCCACTACTACGCCGGACACTCCCTCGGCGAAATCACCGCCGCCCACCTCGC 326
              | || ||||||              |||||  |  |    | || ||  ||| | | |
Db        891 CCTCCCCCCCCC----------------CTCCCCCCCCCCCCCCCCCCCCCTCCAGCCCCC 848

Qy        327 CGGCATCCTCACCCTCACCGACGCCACCACCCTCATCACCCAACGCGCCACCCTCATGCA 386
              |  |  || | ||| |||| || || | |||| | || ||| |  | | ||| |||    |
Db        847 CCCCCCCCACCCCCACACCAACCCCCCTACCCCCCTCCCCCCCCACCCCCCCCCCACCACC 788

Qy        387 AACCATGCCCCCCGGCACCATGACCACCCTCCACACCACCCCCCACCACATCACCCACCA 446
              |  |   || ||| |  ||   || ||| || | || |||| | || |  | ||| ||
Db        787 ACACCCCCCACCCACCCACACCCCCCCCCCCCCCCCCCCCCCACCCCCCCCCCCCCCCC 728

Qy        447 CCTCACCGCCCACGAAAACGACCTCGCCATCGCCGCCATCAACACCCCCACCTCCCTCGT 506
              || | || ||| |    |  || | ||  | || ||  |  | ||||| || ||| |
Db        727 CCCCCCCACCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 668

Qy        507 CATCAGCGGCACCCCCCACACCGTCCAACACATCACCACCCTCTGCCAACAACAAGGCAT 566
              || |  | |||||| |||   ||  | | | || ||| |  ||  |  |    |
Db        667 CAACCACCACCCCCCCCTCACACCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 608

Qy        567 CAAAACCAAAACCCTCCCCACCAACCACGCCTTCCACTCCCCCCACACCAACCCCATCCT 626
              |    ||    ||| |||| |    | | ||  |||| |||||| ||||  ||||  ||
Db        607 CCCCCCCCCCCCCCGCCCCCCACCACCCCCCCCCCCACCCCCCCCCCACCCCCCCCCCCC 548

Qy        627 CAACCAACTCCACCAGCACACCCAAACCCTCACCTACCACCCACCCCACACCCCCCTCAT 686
              |  |  | || ||  | |  ||    || | || |||  ||  ||  | |||||| |
Db        547 CCGCATCCCCCCCCCCCCCCGCCCCACCCCCCCCACCCACCCACCGGCCCCCCCCACGC 488

Qy        687 CACCGCCAACACCCCACCCGACCAACTCCTCACCCCCCACTACTGGACCCAACAAGCCCG 746
              |  | ||  |||||||||| ||  | || | |||||| |        ||| |||   |
Db        487 CCACCCCCCCACCCCACCCCCCCCCCCCCCCCCCCCCCTCCCACCACCCCCACACCNCAC 428

Qy        747 CAACACCGTCGACTACGCCACCACCACCCAAACCCTCCACCAACACGGCGTCACCACCTA 806
              |  | ||  |  |  |  . || || |||   | | ||  ||  | |  |  || | ||
Db        427 CNNCCCCCACCCCGCCCGGCCCCCCCCCCCCTCTCCCCCCCNCCCCCCCCCCANCCCCCT 368

Qy        807 CATCGAACTCGGACCCGACAACACCCTCACCACCCTCACCCACCACAACCTCCCCAACCC 866
              |  |   | |   |||  |  | ||| | || ||| | ||||  | ||  ||     ||
```

```
Db        367 CNNCCCCCCCCCCCCCCCCCCCCCCCCCCGCCGCCCCCCCCCCAAAAAAATTTCTTTCTCCT 308

Qy        867 CCCCACCACCACCCTCACCCTCACCCACCCCCACCACCACCCCCAAACCCACC 919
              | || || || ||| |||||| | |||||| |||||||| |||| ||| ||||
Db        307 CTCCCCCCCCCCCCCCACCCCCCCCCACCACCACCACCNNCCCCCAACACACC 255


RESULT 14
BH146899/c
LOCUS       BH146899               1222 bp    DNA     linear   GSS 27-AUG-2001
DEFINITION  ENTQM47TF Entamoeba histolytica Sheared DNA Entamoeba histolytica
            genomic, DNA sequence.
ACCESSION   BH146899
VERSION     BH146899.1  GI:15302995
KEYWORDS    GSS.
SOURCE      Entamoeba histolytica.
  ORGANISM  Entamoeba histolytica
            Eukaryota; Entamoebidae; Entamoeba.
REFERENCE   1  (bases 1 to 1222)
  AUTHORS   Loftus,B., Wang,Z., Van Aken,S. and Fraser,C.
  TITLE     Determination of clone end sequences from Entamoeba histolytica
            HM1:IMSS sheared DNA library (2001)
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Brendan J Loftus
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0208
            Fax: 301 838 3543
            Email: bjloftus@tigr.org
            Clones are derived from the Entamoeba histolytica HM1:IMSS sheared
            DNA library
            Seq primer: M13-Forward
            Class: shotgun
            High quality sequence start: 452
            High quality sequence stop: 563.
FEATURES             Location/Qualifiers
     source          1. .1222
                     /organism="Entamoeba histolytica"
                     /strain="HM1:IMSS"
                     /db_xref="taxon:5759"
                     /clone_lib="Entamoeba histolytica Sheared DNA"
                     /note="Vector: pHOS1; Site_1: Bst I; Constructed at The
                     Institute for Genomic Research (TIGR), Rockville, MD.
                     Genomic DNA isolated from broth cultures of E. histolytica
                     using a  method described by Clark and Diamond (Clark,
                     C.G., and Diamond, L.S.  (1993) Entamoeba histolytica: a
                     method for isolate identification.  Exp. Parasitol.
                     77:450.). The DNA was mechanically sheared to give a
                     tight size distribution (~2 kb). The v + i method used for
                     the library  construction is described in detail in Smith,
                     H.O. and Venter, J.C.  (Making small insert libraries for
                     whole genome shotgun sequencing  projects. In Genome
                     Sequencing: A Practical Approach, eds. M. Vaudin  and B.
                     Barell, Oxford University Press, 1999)."
BASE COUNT      125 a      0 c   1048 g     49 t
ORIGIN

  Query Match             24.7%;  Score 228.4;  DB 17;  Length 1222;
  Best Local Similarity   53.2%;  Pred. No. 1.6e-27;
  Matches  484;  Conservative    0;  Mismatches  426;  Indels    0;  Gaps    0;

Qy         15 CTCCGGACAGGGCACCCAACGCCCCGGCATGGCCCACGGCCTCTACCACACCCACCCCGT 74
              ||||   |    | ||   | |||   |    | | |  |||   || | |||  ||
Db       1130 CTCCCTTCTCCCCCCCCTCTCCCCCACCCCCCCCTCCCCTCCTTCCCCCCCCCCCTCCTCC 1071

Qy         75 CTTCGCCGCCGCACTCAACGACATCTGCACCCACCTCGACCCCCACCTCGACCACCCCCT 134
              | || || | | |   | ||| | ||  || | ||| || || |  |||| ||
Db       1070 CCCCCCCCCCACTCCCCTCCCCCTCTCCCCCTTCCCCCACCTCCCCCCCCCCCCACTCCTC 1011

Qy        135 CCTCCCCCTCCTCACCCAAAACGACAACGACAACGAGGACGCGGCCGCACTGCTCCAGCA 194
              ||  |||| ||   |||    | | |  | |      | |  || |    | ||  |
Db       1010 CCCTCCCCCCCCTTCCCCCTTCCCCCCCCCCCCCCCCCCCCCCCCCCTCCACCCCCCCCC 951

Qy        195 GACCCGCTACGCCCAGCCCGCCCTCTTCGCCTTCCAGGTCGCCCTCCACCGCCTCCTCAC 254
              ||| |  | |||  ||| ||| |    || ||    | ||| || || || || | |
Db        950 CCCCCCCCTCCCCCCCCCCCCCCCCCCTCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 891

Qy        255 CGACGGCTACCACATCACCCCCCACTACTACGCCGGACACTCCCTCGGCGAAATCACCGC 314
              |  |  |  || |  | ||||| | |    ||   | |  || | |      | || |
Db        890 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCTACCCCCCCCCCTCCCCCCCCCCCCCCCC 831

Qy        315 CGCCCACCTCGCCGGCATCCTCACCCTCACCGACGCCACCACCCTCATCACCCAACGCGC 374
              | ||| || | ||  |  || | ||| | ||  | || || ||||  | |||  | | |
Db        830 CCCCCCCCCCCCCCCCTCCCCCCCCCCCCCCCCCCCCCCCCCCCCTCCCCCCCCTCCCCCC 771

Qy        375 CACCCTCATGCAAACCATGCCCCCCGGCACCATGACCACCCTCCACACCACCCCCCACCA 434
              | ||| |   |  || | ||| ||  |     | ||| || | || |||||  |
Db        770 CCCCCCCCCCCCTCCCCTTCCCTCCCCTCCTCCCCTCCCCCCCCCCCCCCCCCCCCTCC 711

Qy        435 CATCACCCACCACCTCACCGCCCACGAAAACGACCTCGCCATCGCCGCCATCAACACCCC 494
              |  | ||| || || |  | ||| |     | || | || || || ||  |  ||||
Db        710 CCCCCCCCCCCCCCCCCTCCCCTCCTCCCCTCCCCCCCCCTCCCCCCCCCCCCCCCCC 651

Qy        495 CACCTCCCTCGTCATCAGCGGCACCCCCCACACCGTCCAACACATCACCACCCTCTGCCA 554
              |||| ||| |  |  |  | |||||| | ||  ||  | |  |  | ||| |  ||
Db        650 CACCCCCCCCCCCCCCCCCCTCCCCCCCCCCCCCCCCCCCCCCCCCCTCCCCCCCCCCC 591

Qy        555 ACAACAAGGCATCAAAACCAAAACCCTCCCCACCAACCACGCCTTCCACTCCCCCCACAC 614
              |  |   |  |    ||    ||| |||| ||  || | ||| || | |||||| | |
Db        590 CCCTCCCCCCCCCCCCCCCCTCCCCCCCCCCCCCCCCCCCCCTCCCCCCCCCCCCCCC 531

Qy        615 CAACCCCATCCTCAACCAACTCCACCAGCACACCCAAACCCTCACCTACCACCCACCCCA 674
              |  ||||  || |  ||  | || ||  | | |||   ||| | ||  || ||| ||||
Db        530 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCT 471

Qy        675 CACCCCCCTCATCACCGCCAACACCCCACCCGACCAACTCCTCACCCCCCACTACTGGAC 734
              | |||||| |  | || ||    |||| |||  ||  |  | | ||||||     |    |
Db        470 CCCCCCCCCCCCCCCCCCCCCCTCCCCCCCCCCCCCCCCCTCCCCCCCCCCCACCCCCCTC 411

Qy        735 CCAACAAGCCCGCAACACCGTCGACTACGCCACCACCACCCAAACCCTCCACCAACACGG 794
              || |  ||    | ||  |  | | || || || |||   |||||| ||  | |
Db        410 CCCCCCCCCCTCTCCCTCCCCCCCCCCCCCCCCCCCCTCCCCCCCCCTCCCCCCCCCCC 351

Qy        795 CGTCACCACCTACATCGAACTCGGACCCGACAACACCCTCACCACCCTCACCCACCACAA 854
              | | ||  |||   ||     |   ||  |      | | ||||   || || || |
Db        350 CCCCCCCCCCTTTTTCTCTTCCCTCTCCCTCCCTTTTCCCCCCACAAACAACCCCCCCCC 291

Qy        855 CCTCCCCAACCCCCCCACCACCACCCTCACCCTCACCCACCCCCACCACCACCCCCAAAC 914
              || ||||  ||||||| || || ||| | ||| | | | ||||| || || |||||   |
Db        290 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCACCCCCCCCCCCCCCCCCCCCC 231

Qy        915 CCACCTCCTC 924
              || || || |
Db        230 CCCCCCCCCC 221


RESULT 15
BF256588/c
LOCUS       BF256588               1161 bp    mRNA    linear   EST 22-OCT-2001
DEFINITION  HVSMEf0010G19f Hordeum vulgare seedling root EST library HVcDNA0007
            (Etiolated and unstressed) Hordeum vulgare cDNA clone
            HVSMEf0010G19f, mRNA sequence.
ACCESSION   BF256588
VERSION     BF256588.2  GI:16315484
KEYWORDS    EST.
SOURCE      Hordeum vulgare.
  ORGANISM  Hordeum vulgare
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
            ; Triticeae; Hordeum.
REFERENCE   1  (bases 1 to 1161)
  AUTHORS   Wing,R., Close,T.J., Kleinhofs,A., Wise,R., Begum,D., Frisch,D., Yu
            ,Y., Henry,D., Palmer,M., Rambo,T., Simmons,J., Choi,D.W., Fenton
            ,R.D., Oates,R. and Main,D.
  TITLE     Development of a genetically and physically anchored EST resource
```

```
            for barley genomics: Morex unstressed seedling root cDNA library
  JOURNAL   Unpublished (2001)
COMMENT     On Nov 16, 2000 this sequence version replaced gi:11185701.
            Contact: Wing RA
            Clemson University Genomics Institute
            Clemson University
            100 Jordan Hall, Clemson, SC 29634, USA
            Tel: 864 656 7288
            Fax: 864 656 4293
            Email: rwing@clemson.edu
            Total hq bases = 244
            Seq primer: AATTAACCCTCACTAAAGGG
            High quality sequence start: 72
            High quality sequence stop: 1138.
FEATURES             Location/Qualifiers
     source          1. .1161
                     /organism="Hordeum vulgare"
                     /cultivar="Morex"
                     /db_xref="taxon:4513"
                     /clone="HVSMEf0010G19f"
                     /clone_lib="Hordeum vulgare seedling root EST library
                     HVcDNA0007 (Etiolated and unstressed)"
                     /tissue_type="Seedling root"
                     /lab_host="TJC121"
                     /note="Vector: lambdaZAP; Site_1: EcoR1; Site_2: Xho1;
                     Seeds were surface sterilized then germinated under axenic
                     conditions in the dark at room temperature on filter paper
                     with water, nystatin and cefotaxime in covered
                     crystallization dishes. Five-day old seedling roots were
                     then harvested, total RNA was prepared, poly(A) RNA was
                     purified, one primary unamplified cDNA library was made,
                     and 1 million pfu were in vivo excised to give pBluescript
                     SK(-) cDNA phagemids. These steps were performed in the TJ
                     Close laboratory at the University of California,
                     Riverside (Choi, Close, Fenton). Phagemids were plated and
                     picked at the Clemson University Genomics Institute (CUGI)
                     (Begum, Palmer, Frisch, Atkins and Wing). Plasmid DNA
                     preparations, DNA sequencing and sequence analysis were
                     performed at CUGI (Wing, Yu, Frisch, Henry, Simmons, Oates
                     , Rambo, Main). The sequence has been trimmed to remove
                     vector sequence and contains a minimum of 100 bases of
                     phred value 20 or above. For more details on library
                     preparation and sequence analysis see
                     http://www.genome.clemson.edu/projects/barley. To order
                     this clone see http://www.genome.clemson.edu/orders Also
                     see Close TJ, Wing R, Kleinhofs A, Wise R (2001)
                     Genetically and physically anchored EST resources for
                     barley genomics. Barley Genetics Newsletter 31:29-30.
                     (http://wheat.pw.usda.gov/ggpages/bgn/31/cover.html)"
BASE COUNT       83 a     45 c    936 g     58 t     39 others
ORIGIN

  Query Match           24.4%;  Score 225;  DB 12;  Length 1161;
  Best Local Similarity 52.1%;  Pred. No. 5.8e-27;
  Matches  468;  Conservative    0;  Mismatches  430;  Indels    0;  Gaps    0;

Qy         27 CACCCAACGCCCCGGCATGGCCCACGGCCTCTACCACACCCACCCCGTCTTCGCCGCCGC 86
              | |||   |||||  |    ||| |  || |  || | ||| ||||  |  | || || |
Db       1062 CCCCCCCGGCCCCCTCGCCCCCCCCCTCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 1003

Qy         87 ACTCAACGACATCTGCACCCACCTCGACCCCCACCTCGACCACCCCCTCCTCCCCCTCCT 146
              | |  |  |  |  | ||| || |  ||||| || |  || ||||| || ||||| ||
Db       1002 CCCCCCCTTCCCCCCCTCCCCCCCCCCCCCCCCTCCCCCCCCCCCCCCCCCCCCCCCCC 943

Qy        147 CACCCAAAACGACAACGACAACGAGGACGCGGCCGCACTGCTCCAGCAGACCCGCTACGC 206
              | |||    |  |  |  |  |      | |  || | |  | ||  |   ||| |  | |
Db        942 CCCCCCCCCCACCCCCCCCCCCCCCCCCCCTCCCCCCCCCCCCCCCCCCCCCCCCCCCC 883

Qy        207 CCAGCCCGCCCTCTTCGCCTTCCAGGTCGCCCTCCACCGCCTCCTCACCGACGGCTACCA 266
              ||  ||| ||| |  |  |  | |  | ||| ||  | ||| | | ||  | ||   ||
Db        882 CCCCCCCCCCCCCCCCCGCCCCGAACCCCCCCGCCGTCCCCTTCCCGCCCCCCGCCCCCC 823

Qy        267 CATCACCCCCCACTACTACGCCGGACACTCCCTCGGCGAAATCACCGCCGCCCACCTCGC 326
```

```
              | | |||||| |  |  | ||   | | ||| |  |      | || || ||| || | |
Db        822 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCTCCCCCCCCCCCCCCTCCCCCCCCCC 763

Qy        327 CGGCATCCTCACCCTCACCGACGCCACCACCCTCATCACCCAACGCGCCACCCTCATGCA 386
              || |  || | ||| | ||   || || |||  |  | |||  | | || ||| |   |
Db        762 CGCCCCCCCCCCCCCCCCCCTTCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 703

Qy        387 AACCATGCCCCCCGGCACCATGACCACCCTCCACACCACCCCCCACCACATCACCCACCA 446
                |    |||||  | || |  || | | |  | || ||||||  || |  | ||| ||
Db        702 CCCACCCCCCCCCCCCCCCCTCCCCCCACCCACCCCCCCCCCCCCCCCCCCCCCCCCCC 643

Qy        447 CCTCACCGCCCACGAAAACGACCTCGCCATCGCCGCCATCAACACCCCCACCTCCCTCGT 506
              |  | |||||| |       |  || | |   | || ||    || ||||| || ||| |
Db        642 CACCCCCGCCCCCCCCCCCCCCCCGCCCACCCCCCCCCCCACACCCCCCCCCCCCCCCCC 583

Qy        507 CATCAGCGGCACCCCCCACACCGTCCAACACATCACCACCCTCTGCCAACAACAAGGCAT 566
              |  |  |  | ||||||  | ||  | |   |  | || ||| |  ||  ||       |
Db        582 CCCCCCCCCCCCCCCCCCCCCCCCCAACACCCCCCCCCCCCCCCCCCCCCACACCCCCCA 523

Qy        567 CAAAACCAAAACCCTCCCCACCAACCACGCCTTCCACTCCCCCCACACCAACCCCATCCT 626
              |    ||    ||| |||| ||  || | |    |   | ||||   ||  |||   ||
Db        522 CGCCCCCCCCNCCCCCCCCCCCCCCCCCCCGCCGCCGCCGCCCCCTCCCGCCCCTCCCCG 463

Qy        627 CAACCAACTCCACCAGCACACCCAAACCCTCACCTACCACCCACCCCACACCCCCCTCAT 686
              | |||    ||  | |  | |||          | ||  || ||| |||| | ||||||  |
Db        462 CCACCGGGACCCGCCGACCCCCCCCCAGGCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCN 403

Qy        687 CACCGCCAACACCCCACCCGACCAACTCCTCACCCCCCACTACTGGACCCAACAAGCCCG 746
              |          |||| |||  ||  |  | | | | ||    |   || || ||
Db        402 NAAAAGNNNNNCCCCNCCCCCCCCCCCCNCCCCCNCNCCCGCCCCCCCCCAAAAAACTTTT 343

Qy        747 CAACACCGTCGACTACGCCACCACCACCCAAACCCTCCACCAACACGGCGTCACCACCTA 806
                    |    | | || || || |||   ||| || ||  | |  |  || | ||
Db        342 TTCTTTTTTTTTTTCCCCCCCCCCCCCCCCCCCCCCCCCCCNCCCCCCCCACANCCCCCC 283

Qy        807 CATCGAACTCGGACCCGACAACACCCTCACCACCCTCACCCACCACAACCTCCCCAACCC 866
              | |  | |  ||| || | || |  ||| |   |  ||   || ||| ||||  ||
Db        282 CCCCCCANCCCCCCCCAACCANACNCNNACCCCANACCACCNAAACGACCCCCCCCCCCN 223

Qy        867 CCCCACCACCACCCTCACCCTCACCCACCCCCACCACCACCCCCAAACCCACCTCCTC 924
              |||| || || ||| | ||| ||| | |  || ||||   ||| || || |
Db        222 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCNCCCNCCCCCCCCCCNCCCCCCCCCCCCC 165
```

Search completed: June 18, 2003, 01:00:26
Job time : 1187.57 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 17, 2003, 11:26:44 ; Search time 1838.93 Seconds
(without alignments)
15003.005 Million cell updates/sec

Title: US-09-914-286-1_COPY_85_1032
Perfect score: 948
Sequence: 1 gtcttcgtttttcccgggca.....gcaccctggaccacttcctg 948

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		948	100.0	30690	6	E38020	E38020 Avermectin
2		948	100.0	64957	1	AB032367	AB032367 Streptomy
3		946.4	99.8	12381	6	AX006889	AX006889 Sequence
4		928.8	98.0	11096	1	AF275943	AF275943 Streptomy
5		354	37.3	65140	6	AX211705	AX211705 Sequence
6		354	37.3	123580	1	AF263912	AF263912 Streptomy
7		354	37.3	125401	6	AX211739	AX211739 Sequence
8		348.2	36.7	28958	6	AR044578	AR044578 Sequence
9		348.2	36.7	28958	6	I47768	I47768 Sequence 6
10		348.2	36.7	28958	6	I50958	I50958 Sequence 6
11		348.2	36.7	28958	6	I63356	I63356 Sequence 6
12		348.2	36.7	28958	6	I70387	I70387 Sequence 6
13		348.2	36.7	28958	6	I85639	I85639 Sequence 6
14		348.2	36.7	28958	6	I88045	I88045 Sequence 4
15		348.2	36.7	28958	6	I90322	I90322 Sequence 6
16		348.2	36.7	49377	6	I88042	I88042 Sequence 1
17		348.2	36.7	67523	1	SCU24241	U24241 Sorangium c
18		337.4	35.6	41097	1	AF016585	AF016585 Streptomy
19		333	35.1	104326	1	AB070940	AB070940 Streptomy
20		327.4	34.5	113193	1	AF357202	AF357202 Streptomy
21		323.4	34.1	78210	1	AB070949	AB070949 Streptomy
22	c	320.6	33.8	84985	1	SNA278573	AJ278573 Streptomy
23		319.2	33.7	43280	1	SFU78289	U78289 Streptomyce
24		318.4	33.6	20235	1	SERERYAB	M63677 S.erythraea
25		318.4	33.6	20235	6	AR049368	AR049368 Sequence
26		318.4	33.6	20235	6	AR095529	AR095529 Sequence
27		318.4	33.6	20444	1	SEERYABS	X62569 S.erythraea
28		313.6	33.1	31422	6	E38021	E38021 Avermectin
29	c	313.6	33.1	64957	1	AB032367	AB032367 Streptomy
30	c	308.4	32.5	104326	1	AB070940	AB070940 Streptomy
31		305	32.2	47981	1	AF263245	AF263245 Micromono
32		305	32.2	47981	6	AX112026	AX112026 Sequence
33		300.6	31.7	7788	6	AX089456	AX089456 Sequence
34		300.6	31.7	50000	6	AX089416	AX089416 Sequence
35		300.6	31.7	50000	6	AX089420	AX089420 Sequence
36		300.6	31.7	80161	1	AY007564	AY007564 Saccharop
37		300.6	31.7	80161	6	AR165018	AR165018 Sequence
38		296.6	31.3	3185	1	AF411573S4	AF411576 Actinomad
39	c	293.6	31.0	15970	1	AE007061	AE007061 Mycobacte
40	c	293.6	31.0	53450	1	MTV018	AL021899 Mycobacte
41		292.4	30.8	27541	6	AX211706	AX211706 Sequence
42		290	30.6	25150	1	AB070945	AB070945 Streptomy
43		284	30.0	37948	1	AF079138	AF079138 Streptomy
44	c	283.2	29.9	1881	6	AR173226	AR173226 Sequence
45	c	283.2	29.9	27522	1	AB070942	AB070942 Streptomy

ALIGNMENTS

RESULT 1
E38020
LOCUS E38020 30690 bp DNA linear PAT 31-JAN-2002
DEFINITION Avermectin aglycon synthase gene.
ACCESSION E38020
VERSION E38020.1 GI:18626909
KEYWORDS JP 2000245457-A/1.
SOURCE Streptomyces avermitilis.
ORGANISM Streptomyces avermitilis
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE 1 (bases 1 to 30690)
AUTHORS Omura,S. and Ikeda,H.
TITLE Avermectin aglycon synthase gene
JOURNAL Patent: JP 2000245457-A 1 12-SEP-2000;

```
            THE KITASATO INSTITUTE
COMMENT     OS   Streptomyces avermitilis
            PN   JP 2000245457-A/1
            PD   12-SEP-2000
            PF   24-FEB-1999 JP 1999046961
            PR
            PI   SATOSHI OMURA,HARUO IKEDA
            PC   C12N15/00,A61K31/70,C12N1/15,C12N9/88,C12P19/62, PC
            C12Q1/68//C07D493/22,
            PC   (C12N1/15,C12R1:465),(C12N9/88,C12R1:465),C12N15/00 CC
            FH   Key             Location/Qualifiers
            FT   CDS             (1). .(11916)
            FT   CDS             (11971). .(30687).
FEATURES             Location/Qualifiers
     source          1. .30690
                     /organism="Streptomyces avermitilis"
                     /db_xref="taxon:33903"
BASE COUNT     5355 a  12455 c   8617 g   4263 t
ORIGIN

  Query Match             100.0%;  Score 948;  DB 6;  Length 30690;
  Best Local Similarity   100.0%;  Pred. No. 4.2e-99;
  Matches  948;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GTCTTCGTTTTTCCCGGGCAGGGCCCGCAATGGCCGGGCATGGGAAGGGAACTTCTCGAC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         85 GTCTTCGTTTTTCCCGGGCAGGGCCCGCAATGGCCGGGCATGGGAAGGGAACTTCTCGAC 144

Qy         61 GCTTCCGACGTCTTCCGGGAGAGCGTCCGCGCCTGCGAAGCCGCGTTCGCGCCCTACGTC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        145 GCTTCCGACGTCTTCCGGGAGAGCGTCCGCGCCTGCGAAGCCGCGTTCGCGCCCTACGTC 204

Qy        121 GACTGGTCGGTGGAGCAGGTGTTGCGGGACTCGCCGGACGCTCCCGGGCTGGACCGGGTG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        205 GACTGGTCGGTGGAGCAGGTGTTGCGGGACTCGCCGGACGCTCCCGGGCTGGACCGGGTG 264

Qy        181 GACGTCGTCCAGCCGACCCTGTTCGCCGTCATGATCTCCCTGGCCGCCCTCTGGCGCTCG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        265 GACGTCGTCCAGCCGACCCTGTTCGCCGTCATGATCTCCCTGGCCGCCCTCTGGCGCTCG 324

Qy        241 CAAGGGGTCGAGCCGTGCGCGGTGCTGGGACACAGCCTGGGCGAGATCGCGGCAGCCCAC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        325 CAAGGGGTCGAGCCGTGCGCGGTGCTGGGACACAGCCTGGGCGAGATCGCGGCAGCCCAC 384

Qy        301 GTCTCGGGAGGCCTGTCCCTGGCCGACGCCGCACGCGTGGTGACGCTTTGGAGCCAGGCA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        385 GTCTCGGGAGGCCTGTCCCTGGCCGACGCCGCACGCGTGGTGACGCTTTGGAGCCAGGCA 444

Qy        361 CAGACCACCCTTGCCGGGACCGGCGCGCTCGTCTCCGTCGCCGCCACGCCGGATGAGCTC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        445 CAGACCACCCTTGCCGGGACCGGCGCGCTCGTCTCCGTCGCCGCCACGCCGGATGAGCTC 504

Qy        421 CTGCCCCGAATCGCTCCGTGGACCGAGGACAACCCGGCGCGGCTCGCCGTCGCAGCCGTC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        505 CTGCCCCGAATCGCTCCGTGGACCGAGGACAACCCGGCGCGGCTCGCCGTCGCAGCCGTC 564

Qy        481 AACGGACCCCGGAGCACAGTCGTTTCCGGTGCCCGCGAGGCCGTCGCGGACCTGGTGGCC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        565 AACGGACCCCGGAGCACAGTCGTTTCCGGTGCCCGCGAGGCCGTCGCGGACCTGGTGGCC 624

Qy        541 GACCTCACCGCCGCGCAGGTGCGCACGCGCATGATCCCGGTGGACGTTCCCGCCCACTCC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        625 GACCTCACCGCCGCGCAGGTGCGCACGCGCATGATCCCGGTGGACGTTCCCGCCCACTCC 684

Qy        601 CCCCTGATGTACGCCATCGAGGAACGGGTCGTCAGCGGCCTGCTGCCCATCACCCCACGC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        685 CCCCTGATGTACGCCATCGAGGAACGGGTCGTCAGCGGCCTGCTGCCCATCACCCCACGC 744

Qy        661 CCCTCCCGCATCCCCTTCCACTCCTCGGTGACCGGCGGCCGCCTCGACACCCGCGAGCTA 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        745 CCCTCCCGCATCCCCTTCCACTCCTCGGTGACCGGCGGCCGCCTCGACACCCGCGAGCTA 804

Qy        721 GACGCGGCGTACTGGTACCGCAACATGTCGAGCACGGTCCGGTTCGAGCCCGCCGCCCGG 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        805 GACGCGGCGTACTGGTACCGCAACATGTCGAGCACGGTCCGGTTCGAGCCCGCCGCCCGG 864

Qy        781 CTGCTTCTGCAGCAGGGGCCCAAGACGTTCGTCGAGATGAGCCCGCACCCGGTGCTGACC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        865 CTGCTTCTGCAGCAGGGGCCCAAGACGTTCGTCGAGATGAGCCCGCACCCGGTGCTGACC 924

Qy        841 ATGGGCCTCCAGGAGCTCGCCCCGGACCTGGGCGACACCACCGGCACCGCCGACACCGTG 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        925 ATGGGCCTCCAGGAGCTCGCCCCGGACCTGGGCGACACCACCGGCACCGCCGACACCGTG 984

Qy        901 ATCATGGGCACGCTGCGCCGCGGCCAGGGCACCCTGGACCACTTCCTG 948
              ||||||||||||||||||||||||||||||||||||||||||||||||
Db        985 ATCATGGGCACGCTGCGCCGCGGCCAGGGCACCCTGGACCACTTCCTG 1032


RESULT 2
AB032367
LOCUS       AB032367               64957 bp    DNA     linear   BCT 18-SEP-1999
DEFINITION  Streptomyces avermitilis polyketide synthase gene cluster (aveA1,
            aveA2, aveA3, aveA4) and aveC, aveE genes, complete cds.
ACCESSION   AB032367
VERSION     AB032367.1  GI:5902890
KEYWORDS    AveC; cytochrome P450 hydroxylase; type I polyketide synthase AVES
            4; type I polyketide synthase AVES 3; type I polyketide synthase
            AVES 2; type I polyketide synthase AVES 1.
SOURCE      Streptomyces avermitilis DNA.
  ORGANISM  Streptomyces avermitilis
            Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
            Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1  (bases 1 to 64957)
  AUTHORS   Ikeda,H., Nonomiya,T., Usami,M., Ohta,T. and Omura,S.
  TITLE     Organization of the biosynthetic gene cluster for the polyketide
            anthelmintic macrolide avermectin in Streptomyces avermitilis
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9509-9514 (1999)
  MEDLINE   99380548
REFERENCE   2  (bases 1 to 64957)
  AUTHORS   Ikeda,H., Nonomiya,T., Usami,M., Ohta,T. and Omura,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-SEP-1999) Haruo Ikeda, School of Pharmaceutical
            Sciences, Kitasato University, Microbial chemistry; 5-9-1
            Shirokane, Minato-ku, Tokyo 108-8641, Japan
            (E-mail:ikeda@mc.pharm.kitasato-u.ac.jp, Tel:+81-3-5791-6242,
            Fax:+81-3-3444-6197)
FEATURES             Location/Qualifiers
     source          1. .64957
                     /organism="Streptomyces avermitilis"
                     /db_xref="taxon:33903"
     gene            101. .12019
                     /gene="aveA1"
     CDS             101. .12019
                     /gene="aveA1"
                     /note="contains a loading module, Ave module 1 and module
                     2
                     multifunctional polyketide synthase"
                     /codon_start=1
                     /transl_table=11
                     /product="type I polyketide synthase AVES 1"
                     /protein_id="BAA84474.1"
                     /db_xref="GI:5902891"
                     /translation="MQRMDGGEEPRPAAGEVLGVADEADGGVVFVFPGQGPQWPGMGR
                     ELLDASDVFRESVRACEAAFAPYVDWSVEQVLRDSPDAPGLDRVDVVQPTLFAVMISL
                     AALWRSQGVEPCAVLGHSLGEIAAAHVSGGLSLADAARVVTLWSQAQTTLAGTGALVS
                     VAATPDELLPRIAPWTEDNPARLAVAAVNGPRSTVVSGAREAVADLVADLTAAQVRTR
                     MIPVDVPAHSPLMYAIEERVVSGLLPITPRPSRIPFHSSVTGGRLDTRELDAAYWYRN
                     MSSTVRFEPAARLLLQQGPKTFVEMSPHPVLTMGLQELAPDLGDTTGTADTVIMGTLR
                     RGQGTLDHFLTSLAQLRGHGETSATTVLSARLTALSPTQQQSLLLDLVRAHTMAVLND
                     DGNERTASDAGPSASFAHLGFDSVMGVELRNRLSKATGLRLPVTLIFDHTTPAAVAAR
                     LRTAALGHLDEDTAPVPDSPSGHGGTAAADDPIAIIGMACRFPGGVRSPKDLWELAAS
                     GGDAIGPFPTDRGWPTEQRHAQDPTQPGTFYPQGGGFLHDAAHFDAGFFGISPREALA
                     MDPQQRLLLETSWEAFERAGIDPLSVRGSRTGVFAGALSFDYGPRMDTASSEGAADVE
                     GHILTGTTGSVLSGRIAYSFGLEGPAITVDTGCSASLVTLHLACQSLRSGECTLALAG
                     GVSVMSTLGMFIEFSRQRGLSVDGRCKAYSAAADGTGWGEGVGMLLVERLSDAVRLGH
```

RVLAVVRGSAVNQDGASNGLTAPNGPAQERVIRQALANAGLSVADVDVVEGHGTGTTL
GDPIEAQALLATYGQRAGDRPLWLGSLKSNIGHTMAAAGVGGVIKMVMALREGVLPRT
LHVDKPSPQVDWSAGAVRLLTEAVPWPGDAAGRLRRAGVSSFGIGGTNAHVILEEAPA
AGGCVAGGGVLEGAPGLAISVAESVAAPVAVSAPVAESVPVPVPVPVPVSARSEAG
LRAQAEALRQYVAVRPDVSLADVGAGLACGRAVLEHRAVVLAADREELVQGLGALAAG
EPDRRVTTGHAPGGDRGGVVFVFPGQGGQWAGMGVRLLASSPVFARRMQACEEALAPW
VDWSVVDILRRDAGDAVWERADVVQPVLFSVMVSLAALWRSYGIEPDAVLGHSQGEIA
AAHVCGALSLKDAAKTVALRSRALAAVRGRGGMASVPLPAQEVEQLIGERWAGRLWVA
AVNGPRSTAVSGDAEAVDEVLAYCAGTGVRARRIPVDYASHCPHVQPLREELLELLGD
ISPQPSGVPFFSTVEGTWLDTTTLDAAYWYRNLHQPVRFSDAVQALADDGHRVFVEVS
PHPTLVPAIEDTTEDTAEDVTAIGSLRRGDNDTRRFLTALAHTHTTGIGTPTTWHHHY
THHHTHPHPHTHLDLPTYPFQHQHYWLESSQPGAGSGSGAGAGSGAGSGRAGTAGGTA
EVESRFWDAVARQDLETVATTLAVPPSAGLDTVVPALSAWHRHQHDQARINTWTYQET
WKPLTLPTTHQPHQTWLIAIPETQTHHPHITNILTNLHHHGITPIPLTLNHTHTNPQH
LHHTLHHTRQQAQNHTTGAITGLLSLLALDETPHPHHPHTPTGTLLNLTLTQTHTQTH
PPTPLWYATTNATTTHPNDPLTHPTQAQTWGLARTTLLEHPTHTAGIIDLPTTPTPHT
LQHLTQTLTQPHHQTQLAIRTTGTHTRRLTPTTLTPTHQPPTPTPHGTTLITGGTGAL
ATHLTHHLTTHQPTQHLLLTSRTGPHTPHAQHLTTQLQQKGIHLTITTCDTSNPDQLQ
QLLNTIPPQHPLTTVIHTAGILDDATLTNLTPTQLNNVLRAKAHSAHLLHQLTQHTPL
TAPVLYSSAAATFGAPGQANYAAANAYLDALAHHRHTHHLPATSIAWGTWQGNGLADS
DKARAYLDRRGFRPMSPELATAAVTQAIADTERPYVVIADIDWSKIEHTSQTSDLVSA
ARAREPAVQRPTPPAELHKTLAHQTSADQRAALLELVRDHVAAVLRHADPKAIAPDQS
FRALGFDSLTAVEFRNLLIKATGLRLPVSLVFDHPTPAKLAVHLQNQLRGTAAESAPS
AAAVTAEASVTEPIAIVGMACRFPGGVTSADDFWDLISSEQDAIGGFPTDRGWDLDTL
YDPDPDHPGTCYTRNGGFLYDAGHFDAEFFGISPREALAMDPQQRLLLETAWETIEHA
GINPHTLHGTPTGVFTGTNGQDYALRVHNAGQSTDGFALTGTAGSVISGRISYTFGFE
GPAVSVDTACSSSLVALHLACQALRAGECSMALAGGVTVMSSPGAFVEFSRQRGLAAD
GHCKAFSAAADGTGWGEGVGMLLVERLSDAHRNGHRVLAVVRGSAVNQDGASNGLTAP
NGPSQQRVIRQALANAGLSAGDVDAVEAHGTGTTLGDPIEAQALLATYGQDRAGEGPL
WLGSVKSNVGHTQAAAGVAGVIKMVMALRHGLLPRTLHVDEPSPHVDWSAGAVQLLTE
TVPWPGGEGRLRRAGVSSFGVSGTNAHVILEEAPADDVPGGPPAGEGDAGSDDEAAAG
SPGVWPWLVSAKSQPALRAQAQALHAHLTDHPGLDLADVGYTLAHARAVFDHRATLIA
ADRDTFLQALQALAAGEPHPAVIHSSAPGGTGTGEAAGKTAFICSGQGTQRPGMAHGL
YHTHPVFAAALNDICTHLDPHLDHPLLPLLTQNDNDNEDAAALLQQTRYAQPALFAFQ
VALHRLLTDGYHITPHYYAGHSLGEITAAHLAGILTLTDATTLITQRATLMQTMPPGT
MTTLHTTPHHITHHLTAHENDLAIAAAINTPTSLVISGTPHTVQHITTLCQQQGIKTKT
LPTNHAFHSPHTNPILNQLHQHTQTLTYHPPHTPLITANTPPDQLLTPHYWTQQARNT
VDYATTTQTLHQHGVTTYIELGPDNTLTTLTHHNLPNPPTTTLTLTHPHHHPQTHLLT
NLAKTTTTWHPHHYTHHDNQPHTHTHLDLPTYPFQHHHYWLESTQPGAGNVSAAGLDP
TEHPLLGATLELATDGGALLAGRLSLRSHPWLADHAVGGTVLLSGATFLELALHAGTY
VGCDRVDELTLHAPLVVPVDGGVSVQVGVAAADGEGRRLVSVYARGGSACGGGGASGG
VWTCHASGVLVEAAAGGVVVDGLAGVWPPRGAVAVDVDGVRDRLAGAGCVLGPVFSGL
RAVWRDGGDLLAEVCLPEEAWGDAAGFGLHPALLDGVVQPLSVLLPGGTGFGEGAGFG
EGVRVPAVWGGVSLHRAGVTGVRVRVSAVGRGGGREAVSVVVGDEAGVPVASVDRLEL
RPVDMGQLRAVSVSAGRRGSLYAVQWAEVGPVPVCGQAWAWHEDVGESGGGPVPGVVV
LRCPDAGAGGGGGGGGGGVGEVVGGVLGVVQGWLGLERFAGSRLVVVTRGAVVAGPE
DGPVDVVGASVWGLVRSAQAEHPDRFVLLDLDTDTGTDLDTGAGAGWGVDGGRVAAVV
ACGEPQLAVRGERLLAARLKRLESSGDVPAQRSGDTRARRSDVPAQRSGGVPARRSVD
VSGREVLPWLSGGSVLVTGGTGVLGAAVARHLAGVCGVRDLLLVSRRGPDAPGAEGLR
AELAALGAEVRIVACDVGERREVVRLLEGVPAGCPLTGVVHAAGVLDDATIASLTPER
LGTVFAAKVDAALLLDELTRGMELSAFVLFSSAAGILGSAGQGNYAAANAALDALAYR
RRAAGLPGVSLAWGLWEEASGMTGHLAGTDHRRIIRSGLHPMSTPDALALFDAALALD
RPVLLPADLRPAPPLPPLLQDLLPATRRRTTRTTTTGGADNGAQLHARLAGQTHEQQH
TTLLALVRSHIATVLGHTTPDTIPPDRAFRDLGFDSLTAVELRNRLSRTTGLRLPTTL
AFDHPNPTTLTHHLHTQLQPQPDNAVAPVLAELDKLESALSALDKTDSASERVTLRLK
SLMLRWNAPQHPTAESADDDEKFTSATEAEIFKFIDNDLGLS"

gene 12071. .30790
/gene="aveA2"

CDS 12071. .30790
/gene="aveA2"
/note="contains Ave module 3, module 4, module 5 and module 6
multifunctional polyketide synthase"
/codon_start=1
/transl_table=11
/product="type I polyketide synthase AVES 2"
/protein_id="BAA84475.1"
/db_xref="GI:5902892"
/translation="MQLANEAKLLEYLKRVTADLDRTRRRLYEVVEREQEPIAIVGMA
CRYPGGATSPTRLWHLVKSQTDAIGEFPTDRGWNLEQLYDPDPDRSGTSYTRSGGFLY
DAGDFDAAFFELSPREALAMDPQQRLLLETTWETFEQGGIDPRSMRGSRTGVFVGINP
EDYTTGYTHQPSNAVEGYLLTGSAASIASGRISYNFGLEGPAITIDTACSSSLVALHL
ACQALRSGECTMALAGGASVMATPFVFTEFSRQRGLAADGRCKAFSAAADGTGWSEGV
GMLLVERLSDARRNGHRVLAVVRGSAVNQDGASNGLTAPNGRSQVKVIRQALANAHLS
PADVDAVEAHGTGTTLGDPIEAQALVEAYGQDRPNGRPLWLGTLKSNIGHSMAAAGVG
GVIKMVMALRNGLLPRTLHVDEPSPHVDWSAGAVQLLTETVPWPGGEGRLRRAGVSSF
GVSGTNAHVILEEAPAHNIPSDTPADDVPGESAADEDAGSGDEAAAGSPGVWPWLVSA
KSQPALRAQAQALHAHLTDHPGLDLADVGYTLAHARAVFDHRATLIAADRDTFLQALQ
ALAAGEPHPAVIHSSAPGGTGTGEAAGKTAFICSGQGTQRPGMAHGLYHTHPVFAAAL
NDICTHLDPHLDHPLLPLLTQDNDNDNEDAAALLQQTRYAQPALFAFQVALHRLLTDT
LTDGYHITPHYYAGHSLGEITAAHLAGILTLTDATTLITQRATLMQTMPPGTMTTLHT
TPHHITHHLTAHENDLAIAAINTPTSLVISGTPHTVQHITTLCQQQGIKTKTLPTNHA
FHSPHTNPILNQLHQHTQTLTYHPPHTPLITANTPPDQLLTPHYWTQQARNTVDYATT
TQTLHQHGVTTYIELGPDNTLTTLTHDNLPNTPTTTLTLTHPHHHPQTHLLTNLAKTT
TTWHPHHYTHHHNQPHTHTHLDLPTYPFQHHHYWLQPPGKPSDPSPSEGREQATTPST
PLRDVLVGKSPQERDEELLRLVRTHAAAVLGHATPEVIVPNKAFKELGFDSLAAIQLR
NRLLADVDLPLPATLIFDYPTPMALCQPLRAAIVGADTGTTTRLPLTAVPADEPIAIV
GMACRYPGDVRTVDDLWQVVSGGHDAIGGFPTNRGWDLDTLYNPDPDHHGTSYTRSGG
FLYDAGNFDPDFFGISPREALAMDPQQRLLLETAWESIEHACINPDSLRGTPTGVFAG
LTYHDYAARFPTAPAGFEGYLGHGSAGSIASGRVAYALGLEGPALTVDTACSSSLVAL
HLACQALRSGECSMALAGGVTVMSTPAGFVEFSRQRGLAVDGRCKAFSAAADGTGWGE
GVGMLLVERLSDARRLGHRILAVVRGSAVNQDGASNGLTAPNGPSQERVIRLALANAD
LTPADVDAVEAHGTGTTLGDPIEAQALLATYGQDRPGNEPLWLGSMKSNIGHAQAAAG
VGGVIKMVMALRNGLLPRTLHVDEPSPHVDWSAGAVQLLTETVPWPGGEGRLRRAGVS
SFGVSGTNAHVILEEAPAHNIPSDTPADDAPGEAAADDVPGEAAGDDAGTGGEATGPA
AGSPGVWPWLVSAKSQPALRAQAQALHAHLTDHPGLDLADVGYTLAHARAVFDHRATL
IAADRDTFLQALQALAAGEPHPAVIHSSAPGGTGTGEAAGKTAFICSGQGTQRPGMAH
GLYHTHPVFAAALNDICTHLDPHLDHPLLPLLTQDPNTQDTTTLEEAAALLQQTPYAQ
PALFAFQVALHRLLTDGYHITPHYYAGHSLGEITAAHLAGILTLTDATTLITQRATLM
QTMPPGTMTTLHTTPHHITHHLTAHENDLAIAAINTPTSLVISGTPHTVQHITTLCQQ
QGIKTKTLPTKNAFHSPHTNPILNQLHQHTQTLTYHPPHTPLITANTPPDQLLTPHYW
TQQARNTVDYATTTQTLHQHGVTTYIELGPDNTLTTLTHHNLPNTPTTTLTLTHPHHH
PQTHLLTNLAKTTTTWHPHHYTHHHNQPHTHTHLDLPTYPFQHQHYWLESTQPGAGSG
SGSGSGRAGTAGGTAEVESRFWDAVARQDLETVATTLAVPPSAGLDTVVPALSAWHRH
QHDQARINTWTYQETWKPLTLPTTHQPHQTWLIAIPETQTHHPHITNILTNLHHHGIT
PIPLTLNHTHTNPQHLHHTRQQAQNHTTGPITGLLSLLALDETPHPHHPHTPTGTLLN
LTLTQTHTQTHPPTPLWYATTNATTTHPNDPLTHPTQAQTWGLARTTLLEHPTHTAGI
IDLPTTPTPHTLHHLTQTLTQPHHQTQLAIRTTGTHTRRLTPTTLTPTHQPPTPTPHG
TTLITGGTGALATHLTHHLTTHQPTQHLLLTSRTGPHTPHAQHLTTQLQQKGIHLTIT
TCDTSNPDQLQQLLNTIPPQHPLTTVIHTAGILDDATLTNLTPTQLNNVLRAKAHSAH
LLHQLTQHTPLNAFVLYSSAAATFGAPGQANYAAANAYLDALAHHRHTHHLPATSIAW
GTWQGNGLATGQVSEHLRRRGMFAMPPELAVTAVDGAIASGRPSLLVADIDWKKLGPV
LSSKSSVLLEDLPQAQGTEEARSTVEQTESTNLRQLLMGRSRSEQEEELLSLVRIHSA
AVLGRDDSEAIPPGRLFRDLGFDSLAAVELRNHLAAQTELALPTTLVFDYPSPTKLAQ
FLLSEIAEFQPDNSTPLPRPRAELDEPIAIVGMACRFPGGVTSADDFWDLISSEQDAI
GGFPTDRGWDLDTLYDPDPDHPGTCYTRNGGFLYDAGHFDAEFFGISPREALAMDPQQ
RLLLETAWETIEHAGINPHTLHGTPTGVFTGTNGQDHAAHIRQAPSGTEGFVLTGAAT
SIASGRISYILGLEGPAVTLDTACSSSLVALHLACQSLRSGECTMALAGGATVMTTPI
TFTEFARQRGLAPDGRCKAFSAAADGTGWGEGVGMLLVERLSDARRNGHRVLAVVRGS
AVNQDGASNGLTAPNGPSQQRVIRQALANADLTPADVDAVEAHGTGTTLGDPIEAQAI
LATYGQDRPGNGPLWLGSVKSNVGHTQAAAGVAGVIKMVMALRHRTLPPTLHADEPSP
HVDWSAGAVQLLTETVPWPGGEGRPRRAGVSSFGVSGTNAHVILEEAPADDVPGGPPA
DEDAGSGEEAAAGSPGVWPWLVSAKSQPALRAQAQALHAHLTDHPGLDLADVGYTLAH
ARAVFDHRATLIAADRDTFLQALQALAAGEPHPAVIHSSAPGGTGTGEAAGKTAFICS
GQGTQRPGMAHGLYHTHPVFAAALNDICTHLDPHLDHPLLPLLTQNDNDNDNEDAAAL
LQQTPYAQPALFAFQVALHRLLTDGYHITPHYYAGHSLGEITAAHLAGILTLTDATTL
ITQRATLMQTMPPGTMTTLHTTPHHITHHLTAHENDLAIAAINTPTSLVISGTPHTVQ
HITTLCQQQGIKTKTLPTNHAFHSPHTNPILNQLHQHTQTLTYHPPHTPLITANTPPD
QLLTPHYWTQQARNTVDYATTTQTLHQHGVTTYIELGPDNTLTTLTHHNLPNTPTTTL
TLTHPHHHPQTHLLTNLAKTTTTWHPHHYTHHHNQPHTHTHLDLPTYPFQHHHYWLEL
PSAQTSPGQRRSRRSAPDTAESEFWDAVNEEDLQSLAETLDIDASALDTVVPALSAWH
RHQHDQARINTWTYQETWKPLTLPTTHQPHQTWLIAIPETQTHHPHITNILTNLHHHG
ITPIPLTVNHTHTNPQHLHHTLHHTRQQAQNHTTGPITGLLSLLALDETPHPHHPHTP
TGTLLNLTLPQTHTQTHPPTPLWYATTNATTTHPNDPLTHPTQAQTWGLARTTLLEHP
THTAGIIDLPTTPTPHTLHHLTQTLTQPHHQTQLAIRTTGTHTRRLTPTTLTPTHQPP
TPTPHGTTLITGGTGALATHLTHHLTTHQPTQHLLLTSRTGPHTPHAQHLTTQLQQKG
IHLTITTCDTSNPDQLQQLLNTIPPQHPLTTVIHTAGVNLFAPVSETDAESFSSVTAA
KATGAAILHELLLDHETLEHFILFSSGAGAWGSGNQCAYSAANAYLDALATHRQTHGL
PGASIAWGPWAGKGMSAGDAAHGYLEKRGILPMEPRMALAAFHRARAQRPNSNLIIAD
IDWERFVPAFTARRHSPLIEDIPEVRQAAQELEAAASTAKTTTAQPIATSLRERLARL
TSSKQNQVLLGLIRTGICTVLGLRNPEGIEDQRAFRDLGFDSLTSAQFSKELAKETGL
PLPPSLVFDYPTPQECAAHLRTQLVDLDDEEDAALSNALPQVAHRRTVEDEPIAIIGM
ACRFPGGVRSADDLWELLASGKDAIGVFPTDRGWDLDTLYDPDPDHPGTCYTRNGGFL
YGAGHFDAEFFGISPREALAMDPQQRLLLETAWETIEHAGINPHTLHGTPTGVFAGIN
AQDHAAHIRQSRDVETIEGYALTGSSGSVASGRVAYTLGLEGPAVSVDTACSSSLVAL

```
  Query Match           100.0%;  Score 948;  DB 1;  Length 64957;
  Best Local Similarity   100.0%;  Pred. No. 3.3e-99;
  Matches  948;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GTCTTCGTTTTTCCCGGGCAGGGCCCGCAATGGCCGGGCATGGGAAGGGAACTTCTCGAC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        185 GTCTTCGTTTTTCCCGGGCAGGGCCCGCAATGGCCGGGCATGGGAAGGGAACTTCTCGAC 244

Qy         61 GCTTCCGACGTCTTCCGGGAGAGCGTCCGCGCCTGCGAAGCCGCGTTCGCGCCCTACGTC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        245 GCTTCCGACGTCTTCCGGGAGAGCGTCCGCGCCTGCGAAGCCGCGTTCGCGCCCTACGTC 304

Qy        121 GACTGGTCGGTGGAGCAGGTGTTGCGGGACTCGCCGGACGCTCCCGGGCTGGACCGGGTG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        305 GACTGGTCGGTGGAGCAGGTGTTGCGGGACTCGCCGGACGCTCCCGGGCTGGACCGGGTG 364

Qy        181 GACGTCGTCCAGCCGACCCTGTTCGCCGTCATGATCTCCCTGGCCGCCCTCTGGCGCTCG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        365 GACGTCGTCCAGCCGACCCTGTTCGCCGTCATGATCTCCCTGGCCGCCCTCTGGCGCTCG 424

Qy        241 CAAGGGGTCGAGCCGTGCGCGGTGCTGGGACACAGCCTGGGCGAGATCGCGGCAGCCCAC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        425 CAAGGGGTCGAGCCGTGCGCGGTGCTGGGACACAGCCTGGGCGAGATCGCGGCAGCCCAC 484

Qy        301 GTCTCGGGAGGCCTGTCCCTGGCCGACGCCGCACGCGTGGTGACGCTTTGGAGCCAGGCA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        485 GTCTCGGGAGGCCTGTCCCTGGCCGACGCCGCACGCGTGGTGACGCTTTGGAGCCAGGCA 544

Qy        361 CAGACCACCCTTGCCGGGACCGGCGCGCTCGTCTCCGTCGCCGCCACGCCGGATGAGCTC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        545 CAGACCACCCTTGCCGGGACCGGCGCGCTCGTCTCCGTCGCCGCCACGCCGGATGAGCTC 604

Qy        421 CTGCCCCGAATCGCTCCGTGGACCGAGGACAACCCGGCGCGGCTCGCCGTCGCAGCCGTC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        605 CTGCCCCGAATCGCTCCGTGGACCGAGGACAACCCGGCGCGGCTCGCCGTCGCAGCCGTC 664

Qy        481 AACGGACCCCGGAGCACAGTCGTTTCCGGTGCCCGCGAGGCCGTCGCGGACCTGGTGGCC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        665 AACGGACCCCGGAGCACAGTCGTTTCCGGTGCCCGCGAGGCCGTCGCGGACCTGGTGGCC 724

Qy        541 GACCTCACCGCCGCGCAGGTGCGCACGCGCATGATCCCGGTGGACGTTCCCGCCCACTCC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        725 GACCTCACCGCCGCGCAGGTGCGCACGCGCATGATCCCGGTGGACGTTCCCGCCCACTCC 784

Qy        601 CCCCTGATGTACGCCATCGAGGAACGGGTCGTCAGCGGCCTGCTGCCCATCACCCCACGC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        785 CCCCTGATGTACGCCATCGAGGAACGGGTCGTCAGCGGCCTGCTGCCCATCACCCCACGC 844

Qy        661 CCCTCCCGCATCCCCTTCCACTCCTCGGTGACCGGCGGCCGCCTCGACACCCGCGAGCTA 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        845 CCCTCCCGCATCCCCTTCCACTCCTCGGTGACCGGCGGCCGCCTCGACACCCGCGAGCTA 904

Qy        721 GACGCGGCGTACTGGTACCGCAACATGTCGAGCACGGTCCGGTTCGAGCCCGCCGCCCGG 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        905 GACGCGGCGTACTGGTACCGCAACATGTCGAGCACGGTCCGGTTCGAGCCCGCCGCCCGG 964

Qy        781 CTGCTTCTGCAGCAGGGGCCCAAGACGTTCGTCGAGATGAGCCCGCACCCGGTGCTGACC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        965 CTGCTTCTGCAGCAGGGGCCCAAGACGTTCGTCGAGATGAGCCCGCACCCGGTGCTGACC 1024

Qy        841 ATGGGCCTCCAGGAGCTCGCCCCGGACCTGGGCGACACCACCGGCACCGCCGACACCGTG 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1025 ATGGGCCTCCAGGAGCTCGCCCCGGACCTGGGCGACACCACCGGCACCGCCGACACCGTG 1084

Qy        901 ATCATGGGCACGCTGCGCCGCGGCCAGGGCACCCTGGACCACTTCCTG 948
              |||||||||||||||||||||||||||||||||||||||||||||||
Db       1085 ATCATGGGCACGCTGCGCCGCGGCCAGGGCACCCTGGACCACTTCCTG 1132


RESULT 3
AX006889
LOCUS       AX006889               12381 bp    DNA     linear   PAT 06-SEP-2000
DEFINITION  Sequence 1 from Patent WO0001827.
ACCESSION   AX006889
VERSION     AX006889.1  GI:9994904
KEYWORDS    .
SOURCE      Streptomyces avermitilis.
  ORGANISM  Streptomyces avermitilis
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1  (bases 1 to 12381)
  AUTHORS   Kellenberger,J.L., Leadlay,P.F., Staunton,J., Mcarthur,H.A. and
            Stutzman-Engwall,K.J.
  TITLE     Polyketides, their preparation, and materials for use therein
  JOURNAL   Patent: WO 0001827-A 1 13-JAN-2000;
            KELLENBERGER JOHANNES LAURENZ (CH); LEADLAY PETER FRANCIS (GB);
            STAUNTON JAMES (GB); BIOTICA TECH LTD (GB); PFIZER (US); MCARTHUR
            HAMISH ALASTAIR IRVIN (US); STUTZMAN ENGWALL KIM JONELLE (US)
FEATURES             Location/Qualifiers
     source          1. .12381
                     /organism="Streptomyces avermitilis"
                     /db_xref="taxon:33903"
BASE COUNT     1884 a   4561 c   4005 g   1931 t
ORIGIN

  Query Match           99.8%;  Score 946.4;  DB 6;  Length 12381;
  Best Local Similarity   99.9%;  Pred. No. 8.6e-99;
  Matches  947;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 GTCTTCGTTTTTCCCGGGCAGGGCCCGCAATGGCCGGGCATGGGAAGGGAACTTCTCGAC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        902 GTCTTCGTTTTTCCCGGGCAGGGCCCGCAATGGCCGGGCATGGGAAGGGAACTTCTCGAC 961

Qy         61 GCTTCCGACGTCTTCCGGGAGAGCGTCCGCGCCTGCGAAGCCGCGTTCGCGCCCTACGTC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        962 GCTTCCGACGTCTTCCGGGAGAGCGTCCGCGCCTGCGAAGCCGCGTTCGCGCCCTACGTC 1021

Qy        121 GACTGGTCGGTGGAGCAGGTGTTGCGGGACTCGCCGGACGCTCCCGGGCTGGACCGGGTG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1022 GACTGGTCGGTGGAGCAGGTGTTGCGGGACTCGCCGGACGCTCCCGGGCTGGACCGGGTG 1081

Qy        181 GACGTCGTCCAGCCGACCCTGTTCGCCGTCATGATCTCCCTGGCCGCCCTCTGGCGCTCG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1082 GACGTCGTCCAGCCGACCCTGTTCGCCGTCATGATCTCCCTGGCCGCCCTCTGGCGCTCG 1141

Qy        241 CAAGGGGTCGAGCCGTGCGCGGTGCTGGGACACAGCCTGGGCGAGATCGCGGCAGCCCAC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1142 CAAGGGGTCGAGCCGTGCGCGGTGCTGGGACACAGCCTGGGCGAGATCGCGGCAGCCCAC 1201

Qy        301 GTCTCGGGAGGCCTGTCCCTGGCCGACGCCGCACGCGTGGTGACGCTTTGGAGCCAGGCA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1202 GTCTCGGGAGGCCTGTCCCTGGCCGACGCCGCACGCGTGGTGACGCTTTGGAGCCAGGCA 1261

Qy        361 CAGACCACCCTTGCCGGGACCGGCGCGCTCGTCTCCGTCGCCGCCACGCCGGATGAGCTC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1262 CAGACCACCCTTGCCGGGACCGGCGCGCTCGTCTCCGTCGCCGCCACGCCGGATGAGCTC 1321

Qy        421 CTGCCCCGAATCGCTCCGTGGACCGAGGACAACCCGGCGCGGCTCGCCGTCGCAGCCGTC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1322 CTGCCCCGAATCGCTCCGTGGACCGAGGACAACCCGGCGCGGCTCGCCGTCGCAGCCGTC 1381

Qy        481 AACGGACCCCGGAGCACAGTCGTTTCCGGTGCCCGCGAGGCCGTCGCGGACCTGGTGGCC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1382 AACGGACCCCGGAGCACAGTCGTTTCCGGTGCCCGCGAGGCCGTCGCGGACCTGGTGGCC 1441

Qy        541 GACCTCACCGCCGCGCAGGTGCGCACGCGCATGATCCCGGTGGACGTTCCCGCCCACTCC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1442 GACCTCACCGCCGCGCAGGTGCGCACGCGCATGATCCCGGTGGACGTTCCCGCCCACTCC 1501

Qy        601 CCCCTGATGTACGCCATCGAGGAACGGGTCGTCAGCGGCCTGCTGCCCATCACCCCACGC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1502 CCCCTGATGTACGCCATCGAGGAACGGGTCGTCAGCGGCCTGCTGCCCATCACCCCACGC 1561

Qy        661 CCCTCCCGCATCCCCTTCCACTCCTCGGTGACCGGCGGCCGCCTCGACACCCGCGAGCTA 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1562 CCCTCCCGCATCCCCTTCCACTCCTCGGTGACCGGCGGCCGCCTCGACACCCGCGAGCTA 1621
```

```
Qy        721 GACGCGGCGTACTGGTACCGCAACATGTCGAGCACGGTCCGGTTCGAGCCCGCCGCCCGG 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1622 GACGCGGCGTACTGGTACCGCAACATGTCGAGCACGGTCCGGTTCGAGCCCGCCGCCCGG 1681

Qy        781 CTGCTTCTGCAGCAGGGGCCCAAGACGTTCGTCGAGATGAGCCCGCACCCGGTGCTGACC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1682 CTGCTTCTGCAGCAGGGGCCCAAGACGTTCGTCGAGATGAGCCCGCACCCGGTGCTGACC 1741

Qy        841 ATGGGCCTCCAGGAGCTCGCCCCGGACCTGGGCGACACCACCGGCACCGCCGACACCGTG 900
              |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db       1742 ATGGGCCTCCAGGAGCTCGCCGCGGACCTGGGCGACACCACCGGCACCGCCGACACCGTG 1801

Qy        901 ATCATGGGCACGCTGCGCCGCGGCCAGGGCACCCTGGACCACTTCCTG 948
              ||||||||||||||||||||||||||||||||||||||||||||||||
Db       1802 ATCATGGGCACGCTGCGCCGCGGCCAGGGCACCCTGGACCACTTCCTG 1849


RESULT 4
AF275943
LOCUS       AF275943               11096 bp    DNA     linear   BCT 02-SEP-2000
DEFINITION  Streptomyces avermitilis avermectin polyketide synthase gene,
            partial cds.
ACCESSION   AF275943
VERSION     AF275943.1  GI:9964075
KEYWORDS    .
SOURCE      Streptomyces avermitilis.
  ORGANISM  Streptomyces avermitilis
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1  (bases 1 to 11096)
  AUTHORS   Hong,Y.-S. and Lee,J.J.
  TITLE     Targeted Gene Disruption of the avermectin O-methyltransferase gene
            and polyketide synthase gene from Streptomyces avermitilis
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 11096)
  AUTHORS   Hong,Y.-S. and Lee,J.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-JUN-2000) Anticancer Agent Research Laboratory, Korea
            Research Institute of Bioscience and Biotechnology, P.O. Box 116,
            YuSong-Gu, Taejon 305-600, South Korea
FEATURES             Location/Qualifiers
     source          1. .11096
                     /organism="Streptomyces avermitilis"
                     /strain="ATCC31271"
                     /db_xref="ATCC:31271"
                     /db_xref="taxon:33903"
     CDS             220. .>11096
                     /note="Avr1"
                     /codon_start=1
                     /transl_table=11
                     /product="avermectin polyketide synthase"
                     /protein_id="AAG09812.1"
                     /db_xref="GI:9964076"
                     /translation="MADEADGGVVFVFPGQGPQWPGMGRELLDASDVFRESVRACEAE
                     FAPYVDWSVEQVLRDSPDAPGLDRVDVVQRTLFAVMISLAALWRSQGVEPCAVLGHSL
                     GEIAAAHVSGGLSLADAARVGDAWSQAQTTLAGTGALVSVAATPDELLPRIAPWTEDN
                     PARLAVAAVNGPRSTVVSGAREAVADLVADLTAAQVRTRMIPVDVPAHSPLMYAIEER
                     VVSGLLPITPRPSRIPFHSSVTGGRLDTRELDAAYWYRNMSSTVRFEPAARLLLQQGP
                     KTFVEMSPHPVLTMGLQELAADLGDTTGTADTVIMGTLRRGQGTLDHFLTSLAQLRGH
                     GETSATTVLSARLTALSPTQQQSLLLDLVRAHTMAVLNDDGNERTASDAGPSASFAHL
                     GFDSVMGVELRNRLSKATGLRLPVTLILDHTTPAAVAARLRTAALGHLDEDTAPVPDS
                     PSGHGGTAAADDPIAIIGMACRFPGGVRSPKDLWELAASGGDAIGPFPTDRGWPTEQR
                     HAQDPTQPGTFYPQGGGFLHDAAHFDAGFFGISPREALAMDPQQRLLLETSWEAFERA
                     GIDPLSVRGSRTGVFAGALSFDYGPRMDTASSEGAADVEGHILTGTTGSVLSGRSAYS
                     FGLEGPAITVDTGCSASLVTLHLACQSLRSGECTFALAGGVSDVHPGMFIEFSRQCGL
                     SVDGRCKAVSAAADGTGWGEGIGLLAVVRGSAVNQDGASNGLTAPNGPAQERVIRQAL
                     ANAGLSVADVDVVEGHGTGTTLGDPIEAQALLATYGQRAGDRPLWLGSLKSNIGHTMA
                     AAGVGGVIKMVMALREGVLPRTLHVDEPSPQGLAAGAVRLLTEAVPWPGDAAGRLRRA
                     GVSSFGIGGTNAHVILEEAPAAGGCVAGGRVLEGAPGLAISVAESVAAPVAVSAPVAE
                     SVPVPVPVPVPVPVSARSEAGLRAQAEALRQYEAVQPDVSLADVGAGLACRQAVLEHH
                     VVILAACTSSSRSAARTTARSSSTARPQARPAPTSARRIGALAAGSGSAALTTGHAPG
                     GDRGGVVFVFPWQGGQWAGMGVRLLCLLRVFARRMQACEEALAPWVDWSVVDILRRDA
                     GDAVWEQADVVQPVLFSVMVSLAALWRSYGIEPNEVLGHSKDEIAAAHIYGALSLKDA
                     AKTVALPPQEVEQLIGERGGRLWVAAVNGPRSTAVSGDAEAVVEVLAYCAGTGVRPRI
                     PVDYASHCPHVQPLREELLELLGDISPQPYGVPFFSTVEGTWLDTTTLDAAYWYRNLH
                     QPVRFSHDVQALADQGHRVLLEVSPHPTLVPAIEDTTEDTARRRHCDRQPPPRRERHP
                     LLPQRLRLDPYYRHRQTHHVAPLLDPPRHLPPPLDAPRPAHLPFQHQHYWLESSQPGA
                     GSGSGAGAGSGAGSGRAGTAGGTAEVESRFWDAVARQDLETVATTLAVPPSAGLDTVV
                     PALSAWHRHQHDQARINTWTYQETWKPLTLPTTHQPHQTWLIAIPETQTHHPHITNIL
                     TNLHHHGITPIPLPLTTHHTNPQHLHHTLHHTRQQAQNHTTGAITGLLSLLALDETPH
                     PHHPHTPTGTLLNLTLTQTHTQTHPPTPLWYATTNATTTHPNDPLTHPTQAQTWGLAR
                     TTLLEHPTHTAGIIDLPTTPTPHTLHHLTQTLTQPHHQTQLAIRTTGTHTRRLTPTTL
                     TPTHQPPTPTPHGTTLITGGTGALATHLTHHLTTHQPTQHLLLTSRTGPHTPHAQHLT
                     TQLQQKGIHLTITTCDTSTPRPTHNNSLNTIPPQHPVTTVIHTGGILDDATLTNLTPT
                     QLNNVLRAKAHSAHLLHQLTQHTPLTAFVLYSSAAATFGAPGQANYAAANAYLDALAH
                     HRHTHHLPATSIAWGTWQGNGLADSDKARAYLDRRGFRPMSPELATAAVTQAIADTER
                     PYVVIADIDWSKIEHTSQTSDLVSAARAREPAVQRPTPPAELHKTLAHQTSADQRAAL
                     LELVRDHVAAVLRHADPKAIAPDQSFRALGFDSLTAVEFRNLLIKATGLRLPVSLVFD
                     HPTPAKLAVHLQNQLRGTAAESAPSAAAVTAEASVTEPIAIVGMACRFPGGVTSADDF
                     WDLISSEQDAIGGFPTDRGWDLDTLYDPDPDHPGTCYTRNGGFLYDAGHFDAEFFGIS
                     PREALAMDPQQRLLLETAWETIEHAGINPHTLHGTPTGVFTGTNGQDYALRVHNAGQS
                     TDGFALTGTAGSVISGRISYTFGFEGPAVSVDTACSSSLVALHLACQALRAGECSMAL
                     AGGVTVMSSPGAFVEFSRQRGLAADGHCKAFSAAADGTGWGEGVGMLLVERLSDAHRN
                     GHRVLAVVRGSAVNQDGASNGLTAPNGPSQQRVIRQALANAGLSAGDVDAVEAHGTGT
                     TLGDPIEAQALLATYGQDRAGEGPLWLGSVKSNVGHTQAAAGVAGVIKMVMALRHGLL
                     PRTLHVDEPSPHVDWSAGAVQLLTETVPWPGGEGRLRRAGVSSFGVSGTNAHVILEEA
                     PADDVPGGPPAGEGDAGSDDEAAAGSPGVWPWLVSAKSQPALRAQAQALHAHLTDHPG
                     LDLADVGYTLAHARAVFDHRATLIAADRDTFLQALQALAAGEPHPAVIHSSAPGGTGT
                     GEAAGKTAPICSGQGTQRPGMAHGLYHTHPVFAAALNDICTHLDPHLDHPLLPLLTQN
                     DNDNEDAAALLQQTRYAQPALFAFQVALHRLLTDGYHITPHYYAGHSLGEITGAHLAG
                     ILTLTDATTLITQRATLMQTMPPGTMTTLHTTPHHITHHLTAHENDLAIAAINTPTSL
                     VISGTPHTVQHITTLCQQQGIKTKTLPTNHAFHSPHTNPILNQLHQHTQTLTYHPPHT
                     PLITANTPPDQLLTPHYWTQQARNTVDYATTTQTLHQHGVTTYIELGPDNTLTTLTHH
                     NLPNTPTTTLTLTHPHHHPQTHLLTNLAKTTTTWHPHHYTHHAQPTPHPHPLDLPTYP
                     FQHHHYWLESTQPGAGNVSAAGLDPTEHPLLGATLELATDGGALLAGRLSLRSHPWLA
                     DHAVGGTVLLSGATFLELALHAGTYVGCDRVDELTLHAPLVVPVDGGVSVQVGVAAAD
                     GKGRRLVSVYARGGSACGGGASGGVWTCHASGVLVEAAAGGVVVDGLAGVWPPRGAV
                     AVDVDGVRDRLAGAGCVLGPVFSGLRAVWRDGGDLLAEVCLPEEAWGDAAGFGLHPAL
                     LNGVVQPLSVLLPGGTGFGKGAGFGKGVRVPAVWGGVSLHRAGVTGVRVRVSAVGRGG
                     GREAVSVVVGDEAGVPVASVDRLELRPVDMGQLRAVSVSAGRRGSLYAVQWAEVGPVP
                     VCGQAWAWHEDVGESGGGPVPGVVVLRCPDAGAGGGGGGGGGGVGEVVGGVLGVVQG
                     WLGLERFAGSRLVVVTRGAVVAGPEDGPVDVVGASVWGLVRSAQAEHPDRFVLLDLDT
                     DTGTDLDTGAGAGWGVDGGRVAAVVACGEPQLAVRGERLLAARLTRLESSGDVPAQRS
                     GDTRARRSDVPAQRSGGVPARRSVDVSGREVLPWLSGGSVLVTGGTGVLGAAVARHLA
                     GVCGVRDLLLVSRRGPDAPGAEGLRRSWPRGAEVRIVACDVGERREVVRLAGGCSCRV
                     SVFVDFSAVASVTVRLPVASDVRKEAAMAYATVEEFTDYLDPDP"
BASE COUNT     1756 a   4059 c   3528 g   1753 t
ORIGIN

  Query Match           98.0%;  Score 928.8;  DB 1;  Length 11096;
  Best Local Similarity  98.7%;  Pred. No. 8.7e-97;
  Matches  936;  Conservative    0;  Mismatches   12;  Indels    0;  Gaps    0;

Qy          1 GTCTTCGTTTTTCCCGGGCAGGGCCCGCAATGGCCGGGCATGGGAAGGGAACTTCTCGAC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        247 GTCTTCGTTTTTCCCGGGCAGGGCCCGCAATGGCCGGGCATGGGAAGGGAACTTCTCGAC 306

Qy         61 GCTTCCGACGTCTTCCGGGAGAGCGTCCGCGCCTGCGAAGCCGCGTTCGCGCCCTACGTC 120
              |||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db        307 GCTTCCGACGTCTTCCGGGAGAGCGTCCGCGCCTGCGAAGCCGAGTTCGCGCCCTACGTC 366

Qy        121 GACTGGTCGGTGGAGCAGGTGTTGCGGGACTCGCCGGACGCTCCCGGGCTGGACCGGGTG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        367 GACTGGTCGGTGGAGCAGGTGTTGCGGGACTCGCCGGACGCTCCCGGGCTGGACCGGGTG 426

Qy        181 GACGTCGTCCAGCCGACCCTGTTCGCCGTCATGATCTCCCTGGCCGCCCTCTGGCGCTCG 240
              ||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db        427 GACGTCGTCCAGCGCACCCTGTTCGCCGTCATGATCTCCCTGGCCGCCCTCTGGCGCTCG 486

Qy        241 CAAGGGGTCGAGCCGTGCGCGGTGCTGGGACACAGCCTGGGCGAGATCGCGGCAGCCCAC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        487 CAAGGGGTCGAGCCGTGCGCGGTGCTGGGACACAGCCTGGGCGAGATCGCGGCAGCCCAC 546

Qy        301 GTCTCGGGAGGCCTGTCCCTGGCCGACGCCGCACGCGTGGTGACGCTTTGGAGCCAGGCA 360
              |||||||||||||||||||||||||||||||||||| |      ||||||||||||
```

```
Db        547 GTCTCGGGAGGCCTGTCCCTGGCCGACGCCGCACGCGTTGGTGACGCTTGGAGCCAGGCA 606

Qy        361 CAGACCACCCTTGCCGGGACCGGCGCGCTCGTCTCCGTCGCCGCCACGCCGGATGAGCTC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        607 CAGACCACCCTTGCCGGGACCGGCGCGCTCGTCTCCGTCGCCGCCACGCCGGATGAGCTC 666

Qy        421 CTGCCCCGAATCGCTCCGTGGACCGAGGACAACCCGGCGCGGCTCGCCGTCGCAGCCGTC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        667 CTGCCCCGAATCGCTCCGTGGACCGAGGACAACCCGGCGCGGCTCGCCGTCGCAGCCGTC 726

Qy        481 AACGGACCCCGGAGCACAGTCGTTTCCGGTGCCCGCGAGGCCGTCGCGGACCTGGTGGCC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        727 AACGGACCCCGGAGCACAGTCGTTTCCGGTGCCCGCGAGGCCGTCGCGGACCTGGTGGCC 786

Qy        541 GACCTCACCGCCGCGCAGGTGCGCACGCGCATGATCCCGGTGGACGTTCCCGCCCACTCC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        787 GACCTCACCGCCGCGCAGGTGCGCACGCGCATGATCCCGGTGGACGTTCCCGCCCACTCC 846

Qy        601 CCCCTGATGTACGCCATCGAGGAACGGGTCGTCAGCGGCCTGCTGCCCATCACCCCACGC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        847 CCCCTGATGTACGCCATCGAGGAACGGGTCGTCAGCGGCCTGCTGCCCATCACCCCACGC 906

Qy        661 CCCTCCCGCATCCCCTTCCACTCCTCGGTGACCGGCGGCCGCCTCGACACCCGCGAGCTA 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        907 CCCTCCCGCATCCCCTTCCACTCCTCGGTGACCGGCGGCCGCCTCGACACCCGCGAGCTA 966

Qy        721 GACGCGGCGTACTGGTACCGCAACATGTCGAGCACGGTCCGGTTCGAGCCCGCCGCCCGG 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        967 GACGCGGCGTACTGGTACCGCAACATGTCGAGCACGGTCCGGTTCGAGCCCGCCGCCCGG 1026

Qy        781 CTGCTTCTGCAGCAGGGGCCCAAGACGTTCGTCGAGATGAGCCCGCACCCGGTGCTGACC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1027 CTGCTTCTGCAGCAGGGGCCCAAGACGTTCGTCGAGATGAGCCCGCACCCGGTGCTGACC 1086

Qy        841 ATGGGCCTCCAGGAGCTCGCCCCGGACCTGGGCGACACCACCGGCACCGCCGACACCGTG 900
              |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db       1087 ATGGGCCTCCAGGAGCTCGCCGCGGACCTGGGCGACACCACCGGCACCGCCGACACCGTG 1146

Qy        901 ATCATGGGCACGCTGCGCCGCGGCCAGGGCACCCTGGACCACTTCCTG 948
              |||||||||||||||||||||||||||||||||||||||||||||||||
Db       1147 ATCATGGGCACGCTGCGCCGCGGCCAGGGCACCCTGGACCACTTCCTG 1194


RESULT 5
AX211705
LOCUS       AX211705               65140 bp    DNA     linear   PAT 06-SEP-2001
DEFINITION  Sequence 1 from Patent WO0159126.
ACCESSION   AX211705
VERSION     AX211705.1  GI:15523937
KEYWORDS    .
SOURCE      Streptomyces noursei.
  ORGANISM  Streptomyces noursei
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1  (bases 1 to 65140)
  AUTHORS   Zotchev,S.B., Sekurova,O.N., Fjaervik,E., Brautaset,T.,
            Stroem,A.R., Valla,S., Ellingsen,T.E., Sletta,H.V. and
            Gulliksen,O.M.
  TITLE     Gene cluster encoding a nystatin polyketide synthase and its
            manipulation and utility
  JOURNAL   Patent: WO 0159126-A 1 16-AUG-2001;
            Norges Teknisk Naturvitenskapelige Universitet (NO) ; STIFTELSEN
            IND OG TEKNISK FORSKNING VED NORGES TEKNISKE HOGSKOLE (NO) ;
            ALPHARMA AS (NO) ; Sinvent AS (NO) ; Zotchev, Sergey Borisovich
            (NO) ; Sekurova, Olga Nikolayivna (NO) ; Fjaervik, Espen (NO) ;
            Brautaset, Trygve (NO) ; Strom, Arne Reidar (NO) ; Valla, Svein
            (NO)
FEATURES             Location/Qualifiers
     source          1. .65140
                     /organism="Streptomyces noursei"
                     /db_xref="taxon:1971"
                     /note="ATCC 11455"
BASE COUNT     8270 a  25171 c  22273 g   9426 t

ORIGIN

  Query Match           37.3%;  Score 354;  DB 6;  Length 65140;
  Best Local Similarity  62.6%;  Pred. No. 4.8e-32;
  Matches  593;  Conservative    0;  Mismatches  340;  Indels   15;  Gaps    2;

Qy          1 GTCTTCGTTTTTCCCGGGCAGGGCCCGCAATGGCCGGGCATGGGAAGGGAACTTCTCGAC 60
              |||||||| || ||||| |||||  | || |||  ||| |||||  | ||||| |||| |
Db      13291 GTCTTCGTCTTCCCCGGCCAGGGATCCCAGTGGTGGGGAATGGGCCGCGAACTCCTCGCC 13350

Qy         61 GCTTCCGACGTCTTCCGGGAGAGCGTCCGCGCCTGCGAAGCCGCGTTCGCGCCCTACGTC 120
                | ||||| || |||||          ||  | |||||   | ||  |||| || ||||||
Db      13351 ACGTCCGAGGTGTTCCGCACCGCGATCGATGACTGCGCGACGGCCCTCGCCCCGTACGTC 13410

Qy        121 GACTGGTCGGTGGAGCAGGTGTTGCGGGACTCGCCGGACGCTCCCGGGCTGGACCGGGTG 180
              ||||||||| || |  | ||  ||    | |  |  ||| |  ||  |||||| ||||||
Db      13411 GACTGGTCGCTGCACGACGTCCTGGCCGGCGAGGGCGACCCCGCCCTGCTGGAGCGGGTG 13470

Qy        181 GACGTCGTCCAGCCGACCCTGTTCGCCGTCATGATCTCCCTGGCCGCCCTCTGGCGCTCG 240
              |||||  |||||||  | |||||||| | ||| ||   ||| |||| ||||||||||
Db      13471 GACGTGGTCCAGCCCGCGCTGTTCGCCATGATGGTCGGGCTGTCCGCGCTCTGGCGCTCC 13530

Qy        241 CAAGGGGTCGAGCCGTGCGCGGTGCTGGGACACAGCCTGGGCGAGATCGCGGCAGCCCAC 300
              || || ||||  |||   || ||| | || |||   | |||||||||||| || |||  |
Db      13531 CACGGCGTCGTCCCGGCGGCCGTGGTCGGCCACTCGCAGGGCGAGATCGCCGCGGCCTGC 13590

Qy        301 GTCTCGGGAGGCCTGTCCCTGGCCGACGCCGCACGCGTGGTGACGCTTTGGAGCCAGGCA 360
              ||| | |||| |||   |||||||||||||| |||||||| ||||  | |||||||||
Db      13591 GTCGCCGGAGCCCTCAGCCTGGCCGACGCCGCCCGCGTGGTGGCGCTGCGCAGCCAGGCA 13650

Qy        361 CAGACCACCCTTGCCGGGACCGGCGCGCTCGTCTCCGTCGCCGCCACGCCGGATGAGCTC 420
              | | |   || ||||  |||||  | || ||| |||| |  ||  | |||  | || ||
Db      13651 CTGCCGCAACTGTCCGGACGCGGCGGCATGATGTCGGTCTCCGCCCCCGTAGAGCGGGTC 13710

Qy        421 CTGCCCCGAATCGCTCCGTGGACCGAGGACAACCCGGCGCGGCTCGCCGTCGCAGCCGTC 480
              |  |  |||| ||||||           | || |  |||  ||||||| || |||
Db      13711 ACCGCACTCCTCGCCCCGTGG-----------CAGGAGGCGCTGTCCGTCGCCGCGGTC 13758

Qy        481 AACGGACCCCGGAGCACAGTCGTTTCCGGTGCCCGCGAGGCCGTCGCGGACCTGGTGGCC 540
              ||||| |||  |  |  ||||| ||||| | |  ||| ||  |||  | ||||    ||
Db      13759 AACGGCCCCTCGTCCGTGGTCGTCTCCGGCGACACCGACGCGCTCGACGCCCTGCACACC 13818

Qy        541 GACCTCACCGCCGCGCAGGTGCGCACGCGCATGATCCCGGTGGACGTTCCCGCCCACTCC 600
              | |  |  |   |  ||||  | |||| | |  | ||||||    || | |||
Db      13819 GCCTGCCAGGAACAGGGCGTGCGGGCCCGCAAGGTGTCCGTGGACTACGCCTCGCACGGG 13878

Qy        601 CCCCTGATGTACGCCATCGAGGAACGGGTCGTCAGCGGCCTGCTGCCCATCACCCCACGC 660
              |  |   |  ||| ||   ||      ||| | ||| |||   |||  ||  ||| |||
Db      13879 CGGCACGTCGAGGCCGTCCGCGACGAACTCGCCCGCGTCCTCGCGCCGGTCGACCCGCGC 13938

Qy        661 CCCTCCCGCATCCCCTTCCACTCCTCGGTGACCGGCGGCCGCCTCGACACCCGCGAGCTA 720
              || ||    | || ||| |||| |||| ||||||| |||| | |||  |  ||   |
Db      13939 GCCCCCGAGGTGCCGTTCTACTCGACGGTCACCGGCGACCGCGTGGACGACGCCGCCTTC 13998

Qy        721 GACGCGGCGTACTGGTACCGCAACATGTCGAGCACGGTCCGGTTCGAGCCCGCCGCCCGG 780
              |||| || |||||||| |||| |       || |||||  | |||   ||| ||||
Db      13999 GACGGCGCCTACTGGTACACCAACCTCCGCCAGACCGTCCGCATGGAGGAGGCCACCCGC 14058

Qy        781 CTGCTTCTGCAGCAGGGGCCCAAGACGTTCGTCGAGATGAGCCCGCACCCGGTGCTGACC 840
              || ||    || | |   ||| ||||| | ||||||||||||||||  ||
Db      14059 GCCCTCCTCGCCGCCGGACACCGCGTCTTCATCGAGGTCAGCCCGCACCCGGTGCTCGCC 14118

Qy        841 ATGGGCCTCCAGGAGCTCGCCCCGGACCTGGGCGACACCACCGGCACCGCCGACACCGTG 900
                  ||||||||        || | | |  ||  ||||||||    |||   | |||
Db      14119 GCCCCGATCCAGGAGACGCAGGAGGCCGTAGCGGAGGCCACCGGCGGGTCCG---CGGTG 14175

Qy        901 ATCATGGGCACGCTGCGCCGCGGCCAGGGCACCCTGGACCACTTCCTG 948
              || | ||| |||| ||||||| | |||||  || |    | |||||||
Db      14176 GTCCTCGGCTCGCTCCGCCGCGACGAGGGCGGCCCGCGGCGCTTCCTG 14223


RESULT 6
AF263912
```

```
LOCUS       AF263912              123580 bp    DNA     linear   BCT 24-MAY-2000
DEFINITION  Streptomyces noursei ATCC 11455 nystatin biosynthetic gene cluster,
            complete sequence.
ACCESSION   AF263912
VERSION     AF263912.1  GI:8050835
KEYWORDS    .
SOURCE      Streptomyces noursei.
  ORGANISM  Streptomyces noursei
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1  (bases 1 to 123580)
  AUTHORS   Brautaset,T., Sekurova,O.N., Sletta,H., Ellingsen,T.E., StrLm,A.R.,
            Valla,S. and Zotchev,S.B.
  TITLE     Biosynthesis of the polyene antifungal antibiotic nystatin in
            Streptomyces noursei ATCC 11455: analysis of the gene cluster and
            deduction of the biosynthetic pathway
  JOURNAL   Chem. Biol. 7 (6), 395-403 (2000)
  MEDLINE   20334850
   PUBMED   10873841
REFERENCE   2  (bases 1 to 123580)
  AUTHORS   Brautaset,T., Sekurova,O.N., Sletta,H., Ellingsen,T.E., Strom,A.R.,
            Valla,S. and Zotchev,S.B.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-MAY-2000) Unigen, NTNU, O. Kyrres gt. 3, Trondheim
            N-7489, Norway
FEATURES             Location/Qualifiers
     source          1. .123580
                     /organism="Streptomyces noursei"
                     /strain="ATCC 11455"
                     /db_xref="ATCC:11455"
                     /db_xref="taxon:1971"
     gene            complement(46. .783)
                     /gene="nysF"
     CDS             complement(46. .783)
                     /gene="nysF"
                     /function="presumably post-translationally modifies ACP
                     domains on PKS"
                     /note="putative 4'-phosphopantheteine transferase"
                     /codon_start=1
                     /transl_table=11
                     /product="NysF"
                     /protein_id="AAF71762.1"
                     /db_xref="GI:8050836"
                     /translation="MIELILPATVATEAAYDDRPRPGDRLLSSEREVIARAVESRQRE
                     FTTVRHLARRALRRLGHPDRAILPNRRGAPQWPPGIVGSMTHCAGYRAAAVSPAELSA
                     AVSIDAEPNGPLPAGVLNAIALPSERPHLVALAAHRPDVHWDRLLFSAKESVFKAWYP
                     LTQRELDFSEAEIVIDPTQGAFTARLLVPGPLLGGRRVTVFPGRWHSTPALLTTAVHL
                     PAPTPRRDREHRTHLTVNSPLPRPTFG"
     gene            complement(867. .2684)
                     /gene="nysG"
     CDS             complement(867. .2684)
                     /gene="nysG"
                     /function="presumably involved in efflux of nystatin"
                     /note="putative transporter (ABC family)"
                     /codon_start=1
                     /transl_table=11
                     /product="NysG"
                     /protein_id="AAF71763.1"
                     /db_xref="GI:8050837"
                     /translation="MASPDDLEEERTAPRPARRLVGLLRPHRRSVALAVSMGVGGIVL
                     NAFGPLLLGRVTDLIADGVLGGVPGPAPGIDFAAIGRLLLVLLALYVVASLFMLAQGR
                     LVASAVWRTIHELRRDAREKLTRLPLRHFDRQPAGELLSRTTNDIDNLQQTLQQTLAE
                     LITSIFSLLTMLVLMLVISPSLAVVMLLSVPVSALIAARISKRAQPHYAAQWSANGTL
                     NAHVEEVCTGHALIKGFDRRAAAEERFDACNDAVYRAAAKAQFASGAMEPVMMFVANL
                     GYVLVAVIGAWKVINGTLTLGDVQAFILYARQFSQPIVEIASVAGRLQSGIASAQRVF
                     TLLDAPEQAPDPLRPGTPARAEGRVEFTDVSFRYSPDTPLIENLSLTVEPGSTVAIVG
                     PTGAGKTTLGNLLMRFYEPDSGRILLDGTDTATMTRDDLRSRFGLVLQDTWLFGGTIA
                     ENIAYGAPGACRADIEEAARATCADRFIRTLPQGYDTVLDDESGTVSAGEKQLLTVAR
                     AFLARPAVLVLDEATSSVDTRTEVLIQRAMNSLRAGRTSFVIAHRLSTIRDADLIVVM
                     DAGRIVEQGTHDQLLCAQGLYARLHAARTHTPTAGAAAG"
     gene            complement(2662. .4416)
                     /gene="nysH"
     CDS             complement(2662. .4416)
                     /gene="nysH"
                     /function="presumably involved in efflux of nystatin"
                     /note="putative transporter (ABC family)"
                     /codon_start=1
                     /transl_table=11
                     /product="NysH"
                     /protein_id="AAF71764.1"
                     /db_xref="GI:8050838"
                     /translation="MLLRLLRAQLRPYAWATAALVALQLVQILGTLLLPTLGAALIDQ
                     GVVRGDGGRITELGVVMGVVALVQIAAALGAAALAARTATAMGRDLRSALFRRILDFS
                     AREIGRFGTPSLLTRSVNDVQQVQNLAQTGFGIVVCAPLMCLGSVLLALRQDVPLALL
                     LVALVLVVAVCFGLLLARMGTLYARMQLTLDRLGRLLREAITGVRVVRSFVRDDHERA
                     RFAQTNDAFLVVSRRVGRLIATMLPVVLLLMNGFTVALLWTGSHRIDAGRMPIGSLSA
                     LLSYLSLILMSVVMLAFVFLSVPRARVCAGRIAEVLDTGSSVAPPAAPQPVRGPAGRI
                     ELCAAGYRYPGAEEPVLRDVDLTVEPGERIAVLGSTGSGKTTLLNLVLRLADATEGAV
                     RVGGTDVRELTAATLAAAVGFVPQRPYLFSGTVASNLRFGRPDATDEELWEALRVAQA
                     ADFVARMPDGLDAEITQGGGNVSGGQRQRLSLARALLRRPEIYLFDDCFSALDQATDA
                     ALRTALVPYTAGATVITVAQRISAGRDADRIVVLDRGRVVAQGTHDVLLRTSPTYREI
                     ALSQLTEEEAAHGLAGRP"
     gene            4714. .5748
                     /gene="nysDIII"
     CDS             4714. .5748
                     /gene="nysDIII"
                     /function="presumably involved in mycosamine biosynthesis"
                     /note="putative dGDP-mannose-4,6-dehydratase"
                     /codon_start=1
                     /transl_table=11
                     /product="NysDIII"
                     /protein_id="AAF71765.1"
                     /db_xref="GI:8050839"
                     /translation="MSKRALITGITGQDGSYLAEHLLSQGYQVWGLIRGQANPRKSRV
                     SRLASELDFIDGDLMDQGSLVSAVDTVQPDEVYNLGAISFVPMSWQQAELVTEVNGMG
                     VLRMLEAIRMVSGLSTSRTVSPRGQIRFYQASSSEMFGKAAETPQRETTLFHPRSPYG
                     AAKAYGHYITRNYRESFGMYAVSGMLFNHESPRRGQEFVTRKISLAVARIKQGLQDKL
                     ALGNLDAVRDWGYAGDYVRAMHLMLQQDAGDDYVIGTGQMHSVRDAVRIAFEHVGLNW
                     EDYVVIDPDLVRPAEVEVLCADSAKAQDRLGWKPDVDFPTLMRMMVDSDLAQVSRENQ
                     YGDVLLAANW"
     gene            5930. .34363
                     /gene="nysI"
     CDS             5930. .34363
                     /gene="nysI"
                     /function="responsible for condensation steps 9 to 14 in
                     the nystatin polyketide backbone synthesis"
                     /note="polyketide synthase"
                     /codon_start=1
                     /transl_table=11
                     /product="NysI"
                     /protein_id="AAF71766.1"
                     /db_xref="GI:8050840"
                     /translation="MDNEQKLRDYLKLATADLRRTRRRVHKLESAAQEPVAIIGMTCR
                     YPGGVRSPEDLWRMVEAGEHGVTPFPTDRGWDLEALAAAPTASGGFLHDAPDFDADFF
                     GISPREAVAMDPQQRVVLESAWEAFERAGIDPTSVKGSRTGVFIGAMAQDYRVGPADG
                     AEGFQLTGNTGSVLSGRISYTFGTVGPAVTVDTACSSSLVAVHLATQALRAGECTLAL
                     AGGVTIMSGPGTFIEMGRQGGLSADGRCRSFGDTADGTGWAEGVGILVLERLSDAVRN
                     GHEILAVVRGTAVNQDGASNGLTAPNGPSQQQVIQQALVNARLAAGDIDVVEAHGTGT
                     TLGDPVEAQALLATYGQNRPADRPLLLGSVKSNLSHTQAAAGVAGVIKMVMAMRHGTL
                     PRTLHAEEPTHHVDWSQGAVRLLTDTTDWPATGAPRRAAVSSFGISGTNAHTIIEQAP
                     EPQPEDAATAQDDAAGSTPATAPVVPGVVPVLLSGRTPDALRGQAAALRAALDTGRRP
                     DLLDLAHSLATTRAGFEHRAVLLATDHPALTDGLTALADADDPAAAPAWITGTTRAET
                     RLAVLFTGQGAQRLGAGRELAARFPAFATALDAALDAFTPHLDRPLREVLWGTDAALL
                     DRTAYAQPALFAVEVALYRLIESFGVRPDHLAGHSVGEIVAAHLAGVLSLADAATLVA
                     ARGRLMQALPDGGAMIAVQASEADVAPLLAGHEDQVAIAAVNGPSAVVLSGAEATVTA
                     LAEQLAADGRKTRRLRVSHAFHSPLMEPMLDAFRAVVEDLTLQPPLLPVVSNLTGKPA
                     TVAQLTSADYWVDHVRHAVRFADGIDWLARHDTTAFLELGPDGVLSAMAQDCLDAADA
                     DAVTLPALRAGRPEEHTLTTALAGLHVHGATLDWTGCFAGTGARRTDLPTYAFQRRRY
                     WPKALQSGTADLRSVGLGAAHHPLLSAAVSLADAGGTLLTGRLSRQTHPWLADHTVRG
                     TTLLPGTAFLELAVRAGDEVGCDRVEELTLAAPLLLPEQGGVQVQLWIGNPDVSGRRT
                     VNVHARPDTGDDTPWTAHATGVLTTADASRQLPASSEQGGTPLAGDPHPALDAAQWPP
                     AGAEPLPLDGHYDRLADGGFGYGPVFQGLRAAWRGGDVVYAEVELPEAGRSDAEAFGL
                     HPALLDAALHAAPFTGLGERGRGGLPFSWEGVSLHAGGATTLRVRLTPVADDALALTV
                     ADGTGAPVLSVDSLVLRSVATQQLDTAAAVARDALFRLDWTPVQPTATDPGPVALLGA
                     DPFGLLTHAGFADAPAYPDLAALAAADGPVPTTVVLSLAGTGDDAADPARSAHRCAAE
                     ALAAVQTWLDHHERFAAARLVFVTRGATVGRDVAAAAVWGLVRSAQSENPGCFALVDL
```

```
DPDGAVGAAALVAALVSGEPQLAVRGDVLRVARLVRRPLTEVGAGADGTGDGVGDGSG
VSFSGEGAVLVTGGTGGLGAVLARHLVAEYGVRDLLLVSRSGERAVGAGELVAELAGV
GARVRVVACDVTDRAAVVELVGGHAVSAVVHAAGVLDDGMVGALTGERLSAVLRPKVD
AVWHLHEATRGLDLDAFVVFSSLAGVFGSPGQANYAAANAFLDALMTRRRAEGLPGLS
LAWGPWEQSGGMTGTLTDVDAERLARSGVPPLSVAQGLALFDAAVAGTDATCVPVRLD
LPVLRARGEVPPLLRSLIRVRARRAAVAGSATAGNLAQRLRRLDEDGRDEMVLDLVRG
QVALVLGHATGGDVDAGRAFRDLGFDSLTAVELRNRLNTVTGLRLPATLVFDYPTVRH
LATYVLDELLGTDAEVATVQPAAAVAVADDPIVIVGMACRYPGGVSSPEDLWRVLTEGT
DAVSGFPTNRGWDVESLYHPDPDHPGTSYTRSGGFLHEAGEFDPGFFGMSPREALATD
SQQRLLLESSWEAIERAGIDPVSLRGSRTGVFAGVMYSDYSAMLASPEFEGFQGSGSS
PSLASGRVAYTLGLEGPAVTVDTACSSSLVAMHWAMQALRSGECGLALAGGVTVMSTP
AVFVDFARQRGLSPDGRCKAFADAADGVGWSEGVGVLVLERQSDAVRNGHEILAVVRG
SAVNQDGASNGLTAPNGPSQQRVIRQALASGGLTAGDVDVVEAHGTGTTLGDPIEAQA
LLATYGRDREPERPLLLGSVKSNLGHTQAAAGVAGVIKMVLAMRHGVVPRTLHVDAPS
SHVDWSEGAVELLSEQAAWPETGRVRRAGVSSFGISGTNVHVIVEQAPGAKAIAAAGA
ARRTPGAVPVLLSGRGRSALRGQAARLLGHLQARPDAELVDVALSLATTRSRFEQRAA
VVAQDRDQLIASLGALAADRPDPAVVEGEAAGRGRTAVLFTGQGSQRAAMGRELHEVQ
PEFAAAFDAVCAVFDPLLDRPLREVVFAEDGSDEAALLDETGWTQPALFAVEVALFRL
VESWGVRPDFVAGHSIGEIAAAHVAGVLTLEDACRLVAARATLMQALPTGGAMIAIQA
TEDEIAAHLDDTVAIAAVNGPQSVVISGDEEAAETIAATFAERGRKTKRLRVSHAFHS
PRMDGMLDAFRIVAEGLTYRAPRIPLVSDLTGRRADDAEVCTAEYWVRHVREAVRFAD
CVRTLRDAGATTFLELGSDGLLTAMAEDTLGDDHDAELVPMLRAGRAEELAAATALAR
LQVRGVDVDWAAYLAGTGARRTDLPTYAFQHAYYWPQLPTPAAALAAADPADQQLWAA
VERGDARELADILGLGEQDLTPLDSLLPALTSWRRGNQEKHLLDTLRYRVEWTRLSKP
TAPVLDGTWLLVASDATAADQPALLDGLADALGSHGARVRRLLLDDSCADRAVLAERL
ARTADVDAATQVLSVLPLDERDADDCPPLTRGLALTVALVQALADTGAQGRLWTATRG
AVSTNPADPVTHPVQAAAWGLGRGVALEHPRLWGGLVDLPQVFDERAGQRLAGILAVK
DAPDGEDQVALRATGVSGRRLVRHTVEALPTAAEFTATGTVLITGGTGGLGAEVARWL
ARAGAQHLVLTSRRGPDAPGAAELRAELEGYGPSVSVVACDVADRDALAAVLTALPEE
LPLTGVVHTAGVGHYGPLDTLSTAEFAGLTAAKLAGAAHLDALLADRELDFFVLFGSI
AGVWGSGNQSAYGAANAYLDALALHRRARGLAATSVAWGPWAEAGMAADDAVSETLRR
QGLGLLDPAPAMTELRRAVVRQDVTVTVADVDWQRYAPLFTSARPSALIAGLPEVRAL
AADERTEQDATGASEVVTRVRALAEPEQLRLLTDLVRTESATVLGHSSADAVPEGRAF
RDVGFDSLTAVELRKRLGAATGLSLPSTMVFDYPTPLELAQYLRAEILGAVLEVAGPV
ATGGADDEPIAIIGMACRFPGGVSSPEQLWDLVASGTDAISEFPVNRGWQTGHLFDPD
PDRPGTTYSTQGGFLHEADEFDPTFFGISPREALVMDPQQRLLLETTWESFERAGIRP
ETLRSTLTGTFVGSSYQEYGLGAGDGTEGHMVTGSSPSVLSGRLSYVFGLEGPAVTVD
TACSSSLVALHLACQSLRNGESNLAVAGGATIMTTPNPFIAFSRQRALAKDGRCKAFS
DDADGMTLAEGVGVVLVERLSDAQRNGHPVLAVLRGSAINQDGASNGLTAPNGPSQQR
VIRQALANARLAPGDIDALEAHGTGTPLGDPIEAQALFATYGRDRDPESALLLGSVKS
NIGHTQSAAGIASVIKMVMALRHSELPPTLHADAPSSHVDWSAGTVRLLTQARAWPET
GRPRRAAVSSFGISGTNAHVLLEQAPVADTPAEERPAVAPVPIAAGVVPWVVTARSAA
ALRGQAERLLAHAETVGTALPAAGPLDIGLSLVSARARFEHRAVVVPPAGTDPLAALR
AVATDGPSPVVARGVADVEGRTVFVFPGQGSQWVGMGSQLLDESAVFAERIAECAAAL
AEFTDWSLVDVLRGVVGAPSLERVDVVQPASFAVMVSLAALWRSRGVLPDAVVGHSQG
EIAAAVVSGALSLRDGARVVALRSQAIGRALAGRGGMMSVALSVDVLEPRLVEFEGRV
SVAAVNGPRSVVVAGEPEALDALHARLTADDIRARRIAVDYASHSHQVEDLHEELLEV
LAELAPRTSEVPFFSTVTGDWLDTARMDAGYWFRNLRGRVRFADAVADLLAAEYRAFV
EVSSHPVLSMAVQEAIDEAGVPAVAAGTLRRDQGGTDRFLLSAAEVFVRGVDVDWAGL
FEGTGASRIDLPTYAFQHEHLWAVPPAPEAVAAADPDDAAFWTAVEDGDVSALTAALG
TDEDSVAAVLPALTSWRRARRDRSTVDAWRYRVAWKPLGGTLPHPSLTGTWLLVTADG
IDDTDVAGALETYGAEVRRLVLDEECVDRAVLRERLAGAEDVTGIVSVLAAAERTDAV
PGTSLVLGTALTVALIQALGDAEIDAPVWALTRGAVSTGRADELTAPVQAQVTGIGWT
AALEHPQRWGGTLDLPAALDARAAQRLAAVLSGALGSDDQLAIRPSGVFTRRIVRAEA
```

```
Query Match           37.3%;  Score 354;  DB 1;  Length 123580;
Best Local Similarity  62.6%;  Pred. No. 3.9e-32;
Matches  593;  Conservative    0;  Mismatches  340;  Indels   15;  Gaps    2;

Qy          1 GTCTTCGTTTTTCCCGGGCAGGGCCCGCAATGGCCGGGCATGGGAAGGGAACTTCTCGAC 60
              ||||||| || ||||| |||||  | || |||  ||| |||||  | ||||| |||| |
Db      73144 GTCTTCGTCTTCCCCGGCCAGGGATCCCAGTGGTGGGGAATGGGCCGCGAACTCCTCGCC 73203

Qy         61 GCTTCCGACGTCTTCCGGGAGAGCGTCCGCGCCTGCGAAGCCGCGTTCGCGCCCTACGTC 120
                | ||||| || |||||         ||   | |||||   | ||  |||| || ||||||
Db      73204 ACGTCCGAGGTGTTCCGCACCGCGATCGATGACTGCGCGACGGCCCTCGCCCCGTACGTC 73263

Qy        121 GACTGGTCGGTGGAGCAGGTGTTGCGGGACTCGCCGGACGCTCCCGGGCTGGACCGGGTG 180
              |||||||||  || |  | ||  ||   | |  |   |||  |  ||  |||||| ||||||
Db      73264 GACTGGTCGCTGCACGACGTCCTGGCCGGCGAGGGCGACCCCGCCCTGCTGGAGCGGGTG 73323

Qy        181 GACGTCGTCCAGCCGACCCTGTTCGCCGTCATGATCTCCCTGGCCGCCCTCTGGCGCTCG 240
              |||||  ||||||||  | |||||||||  |.|||  ||   |||  ||||  ||||||||||||   .
```

```
Db      73324 GACGTGGTCCAGCCCGCGCTGTTCGCCATGATGGTCGGGCTGTCCGCGCTCTGGCGCTCC 73383

Qy        241 CAAGGGGTCGAGCCGTGCGCGGTGCTGGGACACAGCCTGGGCGAGATCGCGGCAGCCCAC 300
              || || ||||  |||   || ||| | || |||   | ||||||||||||| || |||  |
Db      73384 CACGGCGTCGTCCCGGCGGCCGTGGTCGGCCACTCGCAGGGCGAGATCGCCGCGGCCTGC 73443

Qy        301 GTCTCGGGAGGCCTGTCCCTGGCCGACGCCGCACGCGTGGTGACGCTTTGGAGCCAGGCA 360
              ||| | |||| |||   |||||||||||||||| |||||||| ||||  | |||||||||
Db      73444 GTCGCCGGAGCCCTCAGCCTGGCCGACGCCGCCCGCGTGGTGGCGCTGCGCAGCCAGGCA 73503

Qy        361 CAGACCACCCTTGCCGGGACCGGCGCGCTCGTCTCCGTCGCCGCCACGCCGGATGAGCTC 420
              | | |    ||  ||||   |||||   |  | || ||| ||||| |    ||   | ||
Db      73504 CTGCCGCAACTGTCCGGACGCGGCGGCATGATGTCGGTCTCCGCCCCCGTAGAGCGGGTC 73563

Qy        421 CTGCCCCGAATCGCTCCGTGGACCGAGGACAACCCGGCGCGGCTCGCCGTCGCAGCCGTC 480
                | |  |||| ||||||           | || |  |||  ||||||| || |||
Db      73564 ACCGCACTCCTCGCCCCGTGG-----------CAGGAGGCGCTGTCCGTCGCCGCGGTC 73611

Qy        481 AACGGACCCCGGAGCACAGTCGTTTCCGGTGCCCGCGAGGCCGTCGCGGACCTGGTGGCC 540
              ||||| ||| |  |  ||||| ||||| | |  ||| || |||  | ||||    ||
Db      73612 AACGGCCCCTCGTCCGTGGTCGTCTCCGGCGACACCGACGCGCTCGACGCCCTGCACACC 73671

Qy        541 GACCTCACCGCCGCGCAGGTGCGCACGCGCATGATCCCGGTGGACGTTCCCGCCCACTCC 600
              | |  |   |   |  |||||  | |||| | |  | ||||||    || | |||
Db      73672 GCCTGCCAGGAACAGGGCGTGCGGGCCCGCAAGGTGTCCGTGGACTACGCCTCGCACGGG 73731

Qy        601 CCCCTGATGTACGCCATCGAGGAACGGGTCGTCAGCGGCCTGCTGCCCATCACCCCACGC 660
              |  |   | |||| ||   ||  . |||  | |||  |||   |||  ||  ||| |||
Db      73732 CGGCACGTCGAGGCCGTCCGCGACGAACTCGCCCGCGTCCTCGCGCCGGTCGACCCGCGC 73791

Qy        661 CCCTCCCGCATCCCCTTCCACTCCTCGGTGACCGGCGGCCGCCTCGACACCCGCGAGCTA 720
              || ||    | || ||| ||||  |||| ||||||| ||||  | |||   |  ||   |
Db      73792 GCCCCCGAGGTGCCGTTCTACTCGACGGTCACCGGCGACCGCGTGGACGACGCCGCCTTC 73851

Qy        721 GACGCGGCGTACTGGTACCGCAACATGTCGAGCACGGTCCGGTTCGAGCCCGCCGCCCGG 780
              ||||  || |||||||||  |||| |       || |||||  | |||   ||| ||||
Db      73852 GACGGCGCCTACTGGTACACCAACCTCCGCCAGACCGTCCGCATGGAGGAGGCCACCCGC 73911

Qy        781 CTGCTTCTGCAGCAGGGGCCCAAGACGTTCGTCGAGATGAGCCCGCACCCGGTGCTGACC 840
                || ||       || | |     ||| ||||| | ||||||||||||||||  ||
Db      73912 GCCCTCCTCGCCGCCGGACACCGCGTCTTCATCGAGGTCAGCCCGCACCCGGTGCTCGCC 73971

Qy        841 ATGGGCCTCCAGGAGCTCGCCCCGGACCTGGGCGACACCACCGGCACCGCCGACACCGTG 900
                        |||||||       || | | |  ||  ||||||||    |||   | |||
Db      73972 GCCCCGATCCAGGAGACGCAGGAGGCCGTAGCGGAGGCCACCGGCGGGTCCG---CGGTG 74028

Qy        901 ATCATGGGCACGCTGCGCCGCGGCCAGGGCACCCTGGACCACTTCCTG 948
              || | |||  |||| ||||||| | |||||  || |   | |||||||
Db      74029 GTCCTCGGCTCGCTCCGCCGCGACGAGGGCGGCCCGCGGCGCTTCCTG 74076
```

```
RESULT 7
AX211739
LOCUS       AX211739               125401 bp    DNA     linear   PAT 06-SEP-2001
DEFINITION  Sequence 35 from Patent WO0159126.
ACCESSION   AX211739
VERSION     AX211739.1  GI:15523950
KEYWORDS    .
SOURCE      Streptomyces noursei.
  ORGANISM  Streptomyces noursei
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1  (bases 1 to 125401)
  AUTHORS   Zotchev,S.B., Sekurova,O.N., Fjaervik,E., Brautaset,T.,
            Stroem,A.R., Valla,S., Ellingsen,T.E., Sletta,H.V. and
            Gulliksen,O.M.
  TITLE     Gene cluster encoding a nystatin polyketide synthase and its
            manipulation and utility
  JOURNAL   Patent: WO 0159126-A 35 16-AUG-2001;
            Norges Teknisk Naturvitenskapelige Universitet (NO) ; STIFTELSEN
            IND OG TEKNISK FORSKNING VED NORGES TEKNISKE HOGSKOLE (NO) ;
            ALPHARMA AS (NO) ; Sinvent AS (NO) ; Zotchev, Sergey Borisovich
            (NO) ; Sekurova, Olga Nikolayivna (NO) ; Fjaervik, Espen (NO) ;
```

Brautaset, Trygve (NO) ; Strom, Arne Reidar (NO) ; Valla, Svein (NO)
FEATURES Location/Qualifiers
source 1. .125401
/organism="Streptomyces noursei"
/db_xref="taxon:1971"
/note="ATCC 11455"
BASE COUNT 15664 a 49692 c 42871 g 17174 t
ORIGIN

Query Match 37.3%; Score 354; DB 6; Length 125401;
Best Local Similarity 62.6%; Pred. No. 3.9e-32;
Matches 593; Conservative 0; Mismatches 340; Indels 15; Gaps 2;

```
Qy          1 GTCTTCGTTTTTCCCGGGCAGGGCCCGCAATGGCCGGGCATGGGAAGGGAACTTCTCGAC 60
              |||||||| || ||||| ||||  | || |||  ||| |||||  | ||||| |||| |
Db      73552 GTCTTCGTCTTCCCCGGCCAGGGATCCCAGTGGTGGGGAATGGGCCGCGAACTCCTCGCC 73611

Qy         61 GCTTCCGACGTCTTCCGGGAGAGCGTCCGCGCCTGCGAAGCCGCGTTCGCGCCCTACGTC 120
                | ||||| || ||||     ||   | |||||  | ||  |||| || ||||||
Db      73612 ACGTCCGAGGTGTTCCGCACCGCGATCGATGACTGCGCGACGGCCCTCGCCCCGTACGTC 73671

Qy        121 GACTGGTCGGTGGAGCAGGTGTTGCGGGACTCGCCGGACGCTCCCGGGCTGGACCGGGTG 180
              ||||||||| || |  | ||  ||    | |  |   ||| |  ||  |||||| ||||||
Db      73672 GACTGGTCGCTGCACGACGTCCTGGCCGGCGAGGGCGACCCCGCCCTGCTGGAGCGGGTG 73731

Qy        181 GACGTCGTCCAGCCGACCCTGTTCGCCGTCATGATCTCCCTGGCCGCCCTCTGGCGCTCG 240
              |||||  ||||||||  | |||||||| | ||| ||   |||  |||  ||||||||||
Db      73732 GACGTGGTCCAGCCCGCGCTGTTCGCCATGATGGTCGGGCTGTCCGCGCTCTGGCGCTCC 73791

Qy        241 CAAGGGGTCGAGCCGTGCGCGGTGCTGGGACACAGCCTGGGCGAGATCGCGGCAGCCCAC 300
              || || ||||  |||  || ||| | || |||   | ||||||||||| || |||  |
Db      73792 CACGGCGTCGTCCCGGCGGCCGTGGTCGGCCACTCGCAGGGCGAGATCGCCGCGGCCTGC 73851

Qy        301 GTCTCGGGAGGCCTGTCCCTGGCCGACGCCGCCACGCGTGGTGACGCTTTGGAGCCAGGCA 360
              ||| | |||| |||   |||||||||||||| |||||||| ||||  | |||||||||
Db      73852 GTCGCCGGAGCCCTCAGCCTGGCCGACGCCGCCCGCGTGGTGGCGCTGCGCAGCCAGGCA 73911

Qy        361 CAGACCACCCTTGCCGGGACCGGCGCGCTCGTCTCCGTCGCCGCCACGCCGGATGAGCTC 420
              | | |   ||  ||||  |||||   |  | || ||| |||||  |    ||  | ||
Db      73912 CTGCCGCAACTGTCCGGACGCGGCGGCATGATGTCGGTCTCCGCCCCCGTAGAGCGGGTC 73971

Qy        421 CTGCCCCCGAATCGCTCCGTGGACCGAGGACAACCCGGCGCGGCTCGCCGTCGCAGCCGTC 480
                |  |  |||| ||||||          | || |  |||  ||||||| || |||
Db      73972 ACCGCACTCCTCGCCCCGTGG----------CAGGAGGCGCTGTCCGTCGCCGCGGTC 74019

Qy        481 AACGGACCCCGGAGCACAGTCGTTTCCGGTGCCCGCGAGGCCGTCGCGGACCTGGTGGCC 540
              ||||| |||  |  |  |||||  |||||  | |  |||  ||  |||  | ||||    ||
Db      74020 AACGGCCCCTCGTCCGTGGTCGTCTCCGGCGACACCGACGCGCTCGACGCCCTGCACACC 74079

Qy        541 GACCTCACCGCCGCGCAGGTGCGCACGCGCATGATCCCGGTGGACGTTCCCGCCCACTCC 600
              | |  |   |    |  |||||  | |||| | |  | ||||||    || | |||
Db      74080 GCCTGCCAGGAACAGGGCGTGCGGGCCCGCAAGGTGTCCGTGGACTACGCCTCGCACGGG 74139

Qy        601 CCCCTGATGTACGCCATCGAGGAACGGGTCGTCAGCGGCCTGCTGCCCATCACCCCACGC 660
              | |   |  | ||| ||  ||     ||| | ||| |||   |||  ||  ||| |||
Db      74140 CGGCACGTCGAGGCCGTCCGCGACGAACTCGCCCGCGTCCTCGCGCCGGTCGACCCGCGC 74199

Qy        661 CCCTCCCGCATCCCCTTCCACTCCTCGGTGACCGGCGGCCGCCTCGACACCCGCGAGCTA 720
              || ||    | || ||| ||||  |||| ||||||| |||| | |||  |  ||   |
Db      74200 GCCCCCGAGGTGCCGTTCTACTCGACGGTCACCGGCGACCGCGTGGACGACGCCGCCTTC 74259

Qy        721 GACGCGGCGTACTGGTACCGCAACATGTCGAGCACGGTCCGGTTCGAGCCCGCCGCCCGG 780
              ||||  || |||||||  |||| |       || |||||  | |||   ||| ||||
Db      74260 GACGGCGCCTACTGGTACACCAACCTCCGCCAGACCGTCCGCATGGAGGAGGCCACCCGC 74319

Qy        781 CTGCTTCTGCAGCAGGGGCCCAAGACGTTCGTCGAGATGAGCCCGCACCCGGTGCTGACC 840
              || ||     || | |     ||| |||||  | ||||||||||||||||  ||
Db      74320 GCCCTCCTCGCCGCCGGACACCGCGTCTTCATCGAGGTCAGCCCGCACCCGGTGCTCGCC 74379

Qy        841 ATGGGCCTCCAGGAGCTCGCCCCGGACCTGGGCGACACCACCGGCACCGCCGACACCGTG 900
                |||||||         || | | |  ||  ||||||||     |||   | |||
Db      74380 GCCCCGATCCAGGAGACGCAGGAGGCCGTAGCGGAGGCCACCGGCGGGTCCG---CGGTG 74436
```

```
Qy        901 ATCATGGGCACGCTGCGCCGCGGCCAGGGCACCCTGGACCACTTCCTG 948
              || | ||| |||| ||||||| | |||||  || |    | |||||||
Db      74437 GTCCTCGGCTCGCTCCGCCGCGACGAGGGCGGCCCGCGGCGCTTCCTG 74484


RESULT 8
AR044578
LOCUS       AR044578               28958 bp    DNA     linear   PAT 29-SEP-1999
DEFINITION  Sequence 6 from patent US 5817502.
ACCESSION   AR044578
VERSION     AR044578.1  GI:5966043
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 28958)
  AUTHORS   Ligon,J.M., Hill,D.Steven., Lam,S.Ting., Hammer,P.E., van Pee,K.-H.
            and Kirner,S.
  TITLE     Genes for the synthesis of pyrrolnitrin
  JOURNAL   Patent: US 5817502-A 6 06-OCT-1998;
FEATURES             Location/Qualifiers
     source          1. .28958
                     /organism="unknown"
BASE COUNT     4084 a  11863 c   8356 g   4655 t
ORIGIN

Query Match 36.7%; Score 348.2; DB 6; Length 28958;
Best Local Similarity 62.7%; Pred. No. 2.8e-31;
Matches 564; Conservative 0; Mismatches 323; Indels 12; Gaps 1;

Qy          1 GTCTTCGTTTTTCCCGGGCAGGGCCCGCAATGGCCGGGCATGGGAAGGGAACTTCTCGAC 60
              |||||||| ||||| ||||| ||| |||| |||   || ||||          || ||||||
Db       9940 GTCTTCGTCTTTCCTGGGCAAGGCTCGCAGTGGGAAGGGATGGCCCTCTCCCTGCTCGAC 9999

Qy         61 GCTTCCGACGTCTTCCGGGAGAGCGTCCGCGCCTGCGAAGCCGCGTTCGCGCCCTACGTC 120
                | ||    |||||||| |       ||   || |||||   |||| |||| ||  ||||||
Db      10000 TCCTCGCCGGTCTTCCGCGCTCAGCTCGAAGCATGCGAGCGCGCGCTCGCTCCTCACGTC 10059

Qy        121 GACTGGTCGGTGGAGCAGGTGTTGCGGGACTCGCCGGACGCTCCCGGGCTGGACCGGGTG 180
              || |||   ||    ||  ||||   |    || ||| |||   || ||||| ||
Db      10060 GAGTGGAGCCTGCTCGCCGTCCTGCGCCGCGACGAGGGCGCCCCCTCCCTCGACCGCGTC 10119

Qy        181 GACGTCGTCCAGCCGACCCTGTTCGCCGTCATGATCTCCCTGGCCGCCCTCTGGCGCTCG 240
              |||||||| |||||  |||| || |||||||||| ||||||||||||||||||||||||||
Db      10120 GACGTCGTACAGCCCGCCCTCTTTGCCGTCATGGTCTCCCTGGCCGCCCTCTGGCGCTCG 10179

Qy        241 CAAGGGGTCGAGCCGTGCGCGGTGCTGGGACACAGCCTGGGCGAGATCGCGGCAGCCCAC 300
              |  || |||||||||    ||| ||  | || |||||||| ||||||||||| || |||  |
Db      10180 CTCGGCGTCGAGCCCGCCGCCGTCGTCGGCCACAGCCAGGGCGAGATCGCCGCCGCCTTC 10239

Qy        301 GTCTCGGGAGGCCTGTCCCTGGCCGACGCCGCACGCGTGGTGACGCTTTGGAGCCAGGCA 360
              ||| | || |  || ||||| |  ||||| || ||| | |   | ||  | ||  | ||
Db      10240 GTCGCAGGCGCTCTCTCCCTCGAGGACGCGGCGCGCATCGCCGCCCTGCGCAGGAAAGCG 10299

Qy        361 CAGACCACCCTTGCCGGGACCGGCGCGCTCGTCTCCGTCGCCGCCACGCCGGATGAGCTC 420
              |  |||||| | | ||| | |||||    | | | ||||||     |    |  || |||
Db      10300 CTCACCACCGTCGGCGGCAACGGCGGCATGGCCGCCGTCGAGCTCGGCGCCTCCGACCTC 10359

Qy        421 CTGCCCCGAATCGCTCCGTGGACCGAGGACAACCCGGCGCGGCTCGCCGTCGCAGCCGTC 480
              | | ||   |||||| |||  |||                  ||||| ||  ||| ||||||
Db      10360 CAGACCTACCTCGCTCCCTGGGGCGA----------CAGGCTCTCCACCGCCGCCGTC 10407

Qy        481 AACGGACCCCGGAGCACAGTCGTTTCCGGTGCCCGCGAGGCCGTCGCGGACCTGGTGGCC 540
              ||| | ||| ||   ||  |||| ||||| |  | ||  |||||||  |  ||| | | |
Db      10408 AACAGCCCCAGGGCTACCCTCGTATCCGGCGAGCCCGCCGCCGTCGACGCGCTGCTCGAC 10467

Qy        541 GACCTCACCGCCGCGCAGGTGCGCACGCGCATGATCCCGGTGGACGTTCCCGCCCACTCC 600
              | |||||||||| |  |||||  | | |||| |||||   || |||     || ||||||||
Db      10468 GTCCTCACCGCCACCAAGGTGTTCGCCCGCAAGATCCGCGTCGACTACGCCTCCCACTCC 10527

Qy        601 CCCCTGATGTACGCCATCGAGGAACGGGTCGTCAGCGGCCTGCTGCCCATCACCCCACGC 660
              ||| |||| ||||| || | ||   | ||| |   || ||      |||| | || ||
```

```
Db      10528 GCCCAGATGGACGCCGTCCAAGACGAGCTCGCCGCAGGTCTAGCCAACATCGCTCCTCGG 10587

Qy        661 CCCTCCCGCATCCCCTTCCACTCCTCGGTGACCGGCGGCCGCCTCGACACCCGCGAGCTA 720

Db      10588 ACGTGCGAGCTCCCTCTTTATTCGACCGTCACCGGCACCAGGCTCGACGGCTCCGAGCTC 10647

Qy        721 GACGCGGCGTACTGGTACCGCAACATGTCGAGCACGGTCCGGTTCGAGCCCGCCGCCCGG 780

Db      10648 GACGGCGCGTACTGGTATCGAAACCTCCGGCAAACCGTCCTGTTCTCGAGCGCGACCGAG 10707

Qy        781 CTGCTTCTGCAGCAGGGGCCCAAGACGTTCGTCGAGATGAGCCCGCACCCGGTGCTGACC 840

Db      10708 CGGCTCCTCGACGATGGGCATCGCTTCTCCGTCGAGGTCAGCCCCCATCCCGTGCTCACG 10767

Qy        841 ATGGGCCTCCAGGAGCTCGCCCCGGACCTGGGCGACACCACCGGCACCGCCGACACCGT 899

Db      10768 CTCGCCCTCCGCGAGACCTGCGAGCGCTCACCGCTCGATCCCGTCGTCGTCGGCTCCAT 10826


RESULT 9
I47768
LOCUS       I47768                 28958 bp    DNA     linear   PAT 07-OCT-1997
DEFINITION  Sequence 6 from patent US 5639949.
ACCESSION   I47768
VERSION     I47768.1  GI:2471733
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 28958)
  AUTHORS   Ligon,J.M., Hill,D.Steven., Ryals,J.Andrew., Lam,S.Ting. and
            Hammer,P.E.
  TITLE     Genes for the synthesis of antipathogenic substances
  JOURNAL   Patent: US 5639949-A 6 17-JUN-1997;
FEATURES             Location/Qualifiers
     source          1. .28958
                     /organism="unknown"
BASE COUNT     4084 a  11863 c   8356 g   4655 t
ORIGIN

  Query Match           36.7%;  Score 348.2;  DB 6;  Length 28958;
  Best Local Similarity   62.7%;  Pred. No. 2.8e-31;
  Matches  564;  Conservative    0;  Mismatches  323;  Indels   12;  Gaps    1;

Qy          1 GTCTTCGTTTTTCCCGGGCAGGGCCCGCAATGGCCGGGCATGGGAAGGGAACTTCTCGAC 60

Db       9940 GTCTTCGTCTTTCCTGGGCAAGGCTCGCAGTGGGAAGGGATGGCCCTCTCCCTGCTCGAC 9999

Qy         61 GCTTCCGACGTCTTCCGGGAGAGCGTCCGCGCCTGCGAAGCCGCGTTCGCGCCCTACGTC 120

Db      10000 TCCTCGCCGGTCTTCCGCGCTCAGCTCGAAGCATGCGAGCGCGCGCTCGCTCCTCACGTC 10059

Qy        121 GACTGGTCGGTGGAGCAGGTGTTGCGGGACTCGCCGGACGCTCCCGGGCTGGACCGGGTG 180

Db      10060 GAGTGGAGCCTGCTCGCCGTCCTGCGCCGCGACGAGGGCGCCCCCTCCCTCGACCGCGTC 10119

Qy        181 GACGTCGTCCAGCCGACCCTGTTCGCCGTCATGATCTCCCTGGCCGCCCTCTGGCGCTCG 240

Db      10120 GACGTCGTACAGCCCGCCCTCTTTGCCGTCATGGTCTCCCTGGCCGCCCTCTGGCGCTCG 10179

Qy        241 CAAGGGGTCGAGCCGTGCGCGGTGCTGGGACACAGCCTGGGCGAGATCGCGGCAGCCCAC 300

Db      10180 CTCGGCGTCGAGCCCGCCGCCGTCGTCGGCCACAGCCAGGGCGAGATCGCCGCCGCCTTC 10239

Qy        301 GTCTCGGGAGGCCTGTCCCTGGCCGACGCCGCACGCGTGGTGACGCTTTGGAGCCAGGCA 360

Db      10240 GTCGCAGGCGCTCTCTCCCTCGAGGACGCGGCGCGCATCGCCGCCCTGCGCAGGAAAGCG 10299

Qy        361 CAGACCACCCTTGCCGGGACCGGCGCGCTCGTCTCCGTCGCCGCCACGCCGGATGAGCTC 420

Db      10300 CTCACCACCGTCGGCGGCAACGGCGGCATGGCCGCCGTCGAGCTCGGCGCCTCCGACCTC 10359

Qy        421 CTGCCCCGAATCGCTCCGTGGACCGAGGACAACCCGGCGCGGCTCGCCGTCGCAGCCGTC 480

Db      10360 CAGACCTACCTCGCTCCCTGGGGCGA-----------CAGGCTCTCCACCGCCGCCGTC 10407

Qy        481 AACGGACCCCGGAGCACAGTCGTTTCCGGTGCCCGCGAGGCCGTCGCGGACCTGGTGGCC 540

Db      10408 AACAGCCCCAGGGCTACCCTCGTATCCGGCGAGCCCGCCGCCGTCGACGCGCTGCTCGAC 10467

Qy        541 GACCTCACCGCCGCGCAGGTGCGCACGCGCATGATCCCGGTGGACGTTCCCGCCCACTCC 600

Db      10468 GTCCTCACCGCCACCAAGGTGTTCGCCCGCAAGATCCGCGTCGACTACGCCTCCCACTCC 10527

Qy        601 CCCCTGATGTACGCCATCGAGGAACGGGTCGTCAGCGGCCTGCTGCCCATCACCCCACGC 660

Db      10528 GCCCAGATGGACGCCGTCCAAGACGAGCTCGCCGCAGGTCTAGCCAACATCGCTCCTCGG 10587

Qy        661 CCCTCCCGCATCCCCTTCCACTCCTCGGTGACCGGCGGCCGCCTCGACACCCGCGAGCTA 720

Db      10588 ACGTGCGAGCTCCCTCTTTATTCGACCGTCACCGGCACCAGGCTCGACGGCTCCGAGCTC 10647

Qy        721 GACGCGGCGTACTGGTACCGCAACATGTCGAGCACGGTCCGGTTCGAGCCCGCCGCCCGG 780

Db      10648 GACGGCGCGTACTGGTATCGAAACCTCCGGCAAACCGTCCTGTTCTCGAGCGCGACCGAG 10707

Qy        781 CTGCTTCTGCAGCAGGGGCCCAAGACGTTCGTCGAGATGAGCCCGCACCCGGTGCTGACC 840

Db      10708 CGGCTCCTCGACGATGGGCATCGCTTCTCCGTCGAGGTCAGCCCCCATCCCGTGCTCACG 10767

Qy        841 ATGGGCCTCCAGGAGCTCGCCCCGGACCTGGGCGACACCACCGGCACCGCCGACACCGT 899

Db      10768 CTCGCCCTCCGCGAGACCTGCGAGCGCTCACCGCTCGATCCCGTCGTCGTCGGCTCCAT 10826


RESULT 10
I50958
LOCUS       I50958                 28958 bp    DNA     linear   PAT 07-OCT-1997
DEFINITION  Sequence 6 from patent US 5643774.
ACCESSION   I50958
VERSION     I50958.1  GI:2472661
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 28958)
  AUTHORS   Ligon,J.M., Schupp,T., Beck,J.Joseph., Hill,D.Steven.,
            Ryals,J.Andrew., Gaffney,T.Deane., Lam,S.Ting., Hammer,P.E. and
            Uknes,S.Joseph.
  TITLE     Genes for the synthesis of antipathogenic substance
  JOURNAL   Patent: US 5643774-A 6 01-JUL-1997;
FEATURES             Location/Qualifiers
     source          1. .28958
                     /organism="unknown"
BASE COUNT     4084 a  11863 c   8356 g   4655 t
ORIGIN

  Query Match           36.7%;  Score 348.2;  DB 6;  Length 28958;
  Best Local Similarity   62.7%;  Pred. No. 2.8e-31;
  Matches  564;  Conservative    0;  Mismatches  323;  Indels   12;  Gaps    1;

Qy          1 GTCTTCGTTTTTCCCGGGCAGGGCCCGCAATGGCCGGGCATGGGAAGGGAACTTCTCGAC 60

Db       9940 GTCTTCGTCTTTCCTGGGCAAGGCTCGCAGTGGGAAGGGATGGCCCTCTCCCTGCTCGAC 9999

Qy         61 GCTTCCGACGTCTTCCGGGAGAGCGTCCGCGCCTGCGAAGCCGCGTTCGCGCCCTACGTC 120

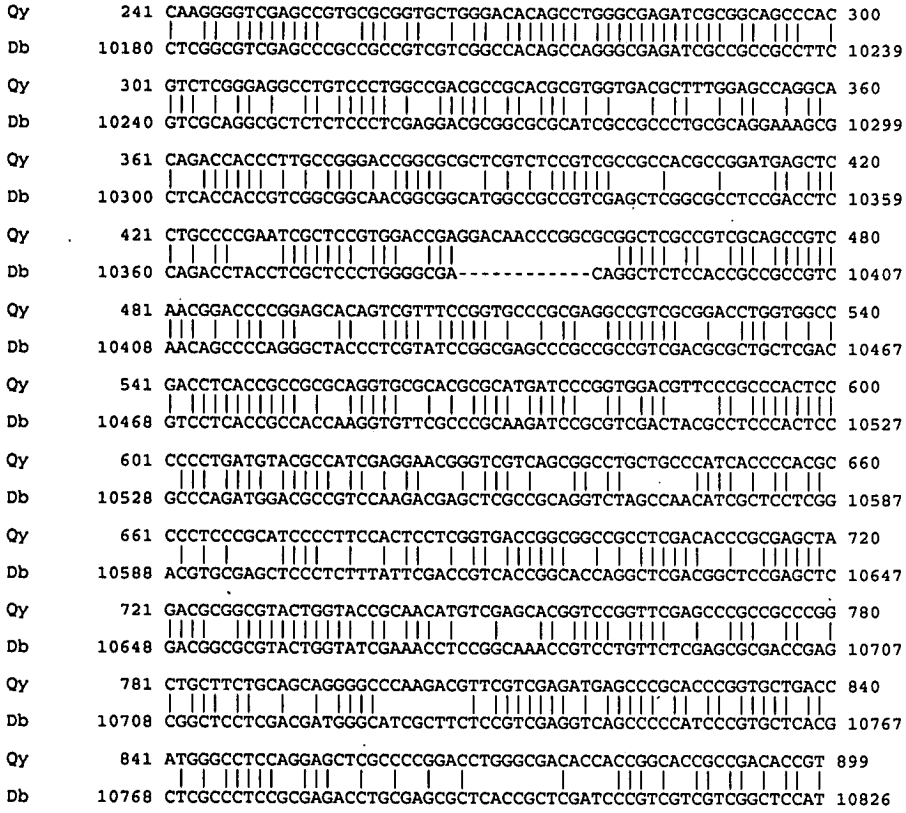
Db      10000 TCCTCGCCGGTCTTCCGCGCTCAGCTCGAAGCATGCGAGCGCGCGCTCGCTCCTCACGTC 10059

Qy        121 GACTGGTCGGTGGAGCAGGTGTTGCGGGACTCGCCGGACGCTCCCGGGCTGGACCGGGTG 180

Db      10060 GAGTGGAGCCTGCTCGCCGTCCTGCGCCGCGACGAGGGCGCCCCCTCCCTCGACCGCGTC 10119

Qy        181 GACGTCGTCCAGCCGACCCTGTTCGCCGTCATGATCTCCCTGGCCGCCCTCTGGCGCTCG 240

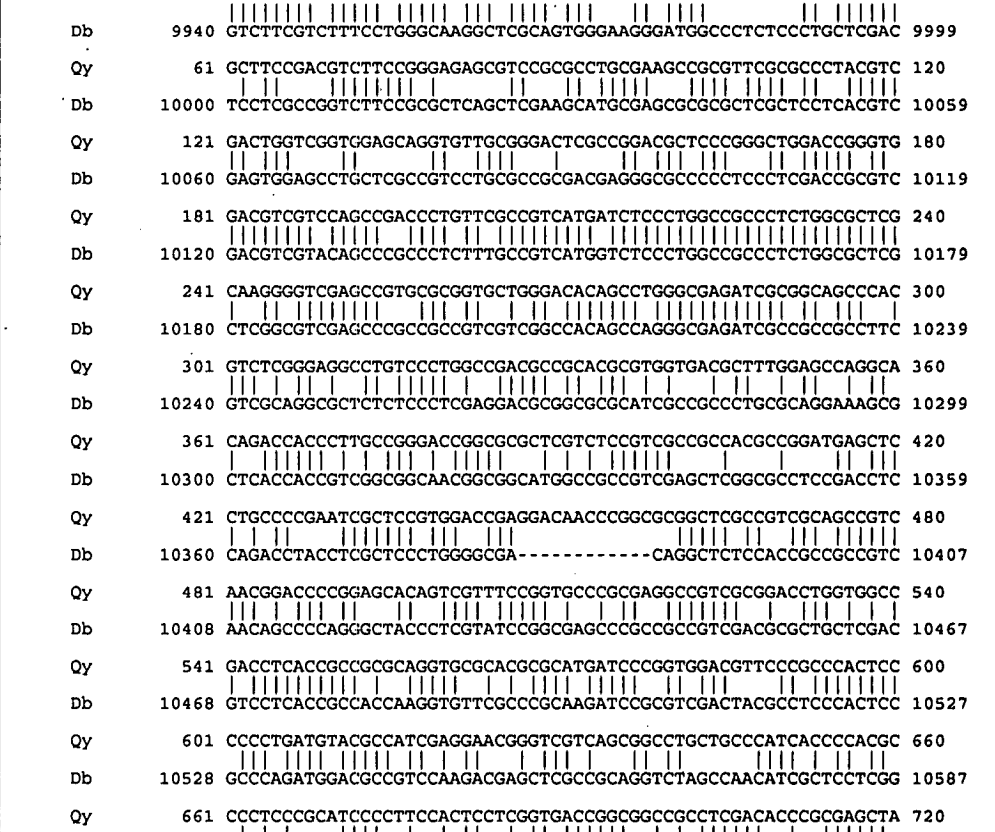
Db      10120 GACGTCGTACAGCCCGCCCTCTTTGCCGTCATGGTCTCCCTGGCCGCCCTCTGGCGCTCG 10179
```



RESULT 11
I63356
LOCUS I63356 28958 bp DNA linear PAT 07-OCT-1997
DEFINITION Sequence 6 from patent US 5662898.
ACCESSION I63356
VERSION I63356.1 GI:2480929
KEYWORDS .
SOURCE Unknown.
ORGANISM Unknown.
Unclassified.
REFERENCE 1 (bases 1 to 28958)
AUTHORS Ligon,J.M., Schupp,T., Beck,J.Joseph., Hill,D.Steven., Ryals,J.Andrew., Gaffney,T.Deane., Lam,S.Ting., Hammer,P.E. and Uknes,S.Joseph.
TITLE Genes for the synthesis of antipathogenic substances
JOURNAL Patent: US 5662898-A 6 02-SEP-1997;
FEATURES Location/Qualifiers
source 1. .28958
/organism="unknown"
BASE COUNT 4084 a 11863 c 8356 g 4655 t
ORIGIN

Query Match 36.7%; Score 348.2; DB 6; Length 28958;
Best Local Similarity 62.7%; Pred. No. 2.8e-31;
Matches 564; Conservative 0; Mismatches 323; Indels 12; Gaps 1;





RESULT 12
I70387
LOCUS I70387 28958 bp DNA linear PAT 03-APR-1998
DEFINITION Sequence 6 from patent US 5679560.
ACCESSION I70387
VERSION I70387.1 GI:3006522
KEYWORDS .
SOURCE Unknown.
ORGANISM Unknown.
Unclassified.
REFERENCE 1 (bases 1 to 28958)
AUTHORS Ligon,J.M., Hill,D.Steven., Lam,S.Ting. and Hammer,P.E.
TITLE Genes for the synthesis of antipathogenic substances

```
  JOURNAL   Patent: US 5679560-A 6 21-OCT-1997;
FEATURES             Location/Qualifiers
     source          1. .28958
                     /organism="unknown"
BASE COUNT     4084 a  11863 c   8356 g   4655 t
ORIGIN

  Query Match            36.7%;  Score 348.2;  DB 6;  Length 28958;
  Best Local Similarity  62.7%;  Pred. No. 2.8e-31;
  Matches  564;  Conservative    0;  Mismatches  323;  Indels   12;  Gaps    1;

Qy          1 GTCTTCGTTTTTCCCGGGCAGGGCCCGCAATGGCCGGGCATGGGAAGGGAACTTCTCGAC 60
              |||||||| ||||| ||||| ||| |||| |||   || ||||         || ||||||
Db       9940 GTCTTCGTCTTTCCTGGGCAAGGCTCGCAGTGGGAAGGGATGGCCCTCTCCCTGCTCGAC 9999

Qy         61 GCTTCCGACGTCTTCCGGGAGAGCGTCCGCGCCTGCGAAGCCGCGTTCGCGCCCTACGTC 120
              | ||    |||||||| |       ||   || |||||   |||| |||| ||  |||||
Db      10000 TCCTCGCCGGTCTTCCGCGCTCAGCTCGAAGCATGCGAGCGCGCGCTCGCTCCTCACGTC 10059

Qy        121 GACTGGTCGGTGGAGCAGGTGTTGCGGGACTCGCCGGACGCTCCCGGGCTGGACCGGGTG 180
              || |||    ||      ||  ||||  |     || ||| |||   || |||||  ||
Db      10060 GAGTGGAGCCTGCTCGCCGTCCTGCGCCGCGACGAGGGCGCCCCCTCCCTCGACCGCGTC 10119

Qy        181 GACGTCGTCCAGCCGACCCTGTTCGCCGTCATGATCTCCCTGGCCGCCCTCTGGCGCTCG 240
              |||||||| |||||  |||| || |||||||| ||||||||||||||||||||||||||||
Db      10120 GACGTCGTACAGCCCGCCCTCTTTGCCGTCATGGTCTCCCTGGCCGCCCTCTGGCGCTCG 10179

Qy        241 CAAGGGGTCGAGCCGTGCGCGGTGCTGGGACACAGCCTGGGCGAGATCGCGGCAGCCCAC 300
              |  || ||||||||   ||| ||  | || |||||||| ||||||||||| || |||  |
Db      10180 CTCGGCGTCGAGCCCGCCGCCGTCGTCGGCCACAGCCAGGGCGAGATCGCCGCCGCCTTC 10239

Qy        301 GTCTCGGGAGGCCTGTCCCTGGCCGACGCCGCACGCGTGGTGACGCTTTGGAGCCAGGCA 360
              ||| | || |  || |||||  |  ||||| || ||| | |   | ||  | ||  | ||
Db      10240 GTCGCAGGCGCTCTCTCCCTCGAGGACGCGGCGCGCATCGCCGCCCTGCGCAGGAAAGCG 10299

Qy        361 CAGACCACCCTTGCCGGGACCGGCGCGCTCGTCTCCGTCGCCGCCACGCCGGATGAGCTC 420
              |  |||||| | | ||| | |||||    | | | ||||||     |    |   || |||
Db      10300 CTCACCACCGTCGGCGGCAACGGCGGCATGGCCGCCGTCGAGCTCGGCGCCTCCGACCTC 10359

Qy        421 CTGCCCCGAATCGCTCCGTGGACCGAGGACAACCCGGCGCGGCTCGCCGTCGCAGCCGTC 480
              | | ||    |||||| |||  |||                ||||| ||  ||| ||||||
Db      10360 CAGACCTACCTCGCTCCCTGGGGCGA------------CAGGCTCTCCACCGCCGCCGTC 10407

Qy        481 AACGGACCCCGGAGCACAGTCGTTTCCGGTGCCCGCGAGGCCGTCGCGGACCTGGTGGCC 540
              ||| | ||| ||   ||  |||| ||||| |  | ||  |||||||  |  ||| | | |
Db      10408 AACAGCCCCAGGGCTACCCTCGTATCCGGCGAGCCCGCCGCCGTCGACGCGCTGCTCGAC 10467

Qy        541 GACCTCACCGCCGCGCAGGTGCGCACGCGCATGATCCCGGTGGACGTTCCCGCCCACTCC 600
              | |||||||||| |  |||||  | | |||| |||||  || |||    || ||||||||
Db      10468 GTCCTCACCGCCACCAAGGTGTTCGCCCGCAAGATCCGCGTCGACTACGCCTCCCACTCC 10527

Qy        601 CCCCTGATGTACGCCATCGAGGAACGGGTCGTCAGCGGCCTGCTGCCCATCACCCCACGC 660
              |||  |||| ||||| || | ||  | ||| |   || ||        |||| | || ||
Db      10528 GCCCAGATGGACGCCGTCCAAGACGAGCTCGCCGCAGGTCTAGCCAACATCGCTCCTCGG 10587

Qy        661 CCCTCCCGCATCCCCTTCCACTCCTCGGTGACCGGCGGCCGCCTCGACACCCGCGAGCTA 720
              | | |   |||  |  | ||  | || ||||||  | | ||||||  |  ||||||
Db      10588 ACGTGCGAGCTCCCTCTTTATTCGACCGTCACCGGCACCAGGCTCGACGGCTCCGAGCTC 10647

Qy        721 GACGCGGCGTACTGGTACCGCAACATGTCGAGCACGGTCCGGTTCGAGCCCGCCGCCCGG 780
              ||||  |||||||||||| || ||| |   |   || |||| ||||  |  |||  ||  |
Db      10648 GACGGCGCGTACTGGTATCGAAACCTCCGGCAAACCGTCCTGTTCTCGAGCGCGACCGAG 10707

Qy        781 CTGCTTCTGCAGCAGGGGCCCAAGACGTTCGTCGAGATGAGCCCGCACCCGGTGCTGACC 840
              | ||| ||  |  | ||||         | ||||||| | ||||| || || ||||| ||
Db      10708 CGGCTCCTCGACGATGGGCATCGCTTCTCCGTCGAGGTCAGCCCCCATCCCGTGCTCACG 10767

Qy        841 ATGGGCCTCCAGGAGCTCGCCCCGGACCTGGGCGACACCACCGGCACCGCCGACACCGT 899
              | | |||||  |||  |  |  |       |  ||| |  || || | || |
Db      10768 CTCGCCCTCCGCGAGACCTGCGAGCGCTCACCGCTCGATCCCGTCGTCGTCGGCTCCAT 10826


RESULT 13
```

```
I85639
LOCUS       I85639                 28958 bp    DNA     linear   PAT 10-JUN-1998
DEFINITION  Sequence 6 from patent US 5698425.
ACCESSION   I85639
VERSION     I85639.1  GI:3205357
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 28958)
  AUTHORS   Ligon,J.M., Schupp,T., Beck,J.Joseph., Hill,D.Steven.,
            Ryals,J.Andrew., Lam,S.Ting., Hammer,P.E. and Uknes,S.Joseph.
  TITLE     Method of protecting plants by transformation with genes for the
            synthesis of antipathogenic substances
  JOURNAL   Patent: US 5698425-A 6 16-DEC-1997;
FEATURES             Location/Qualifiers
     source          1. .28958
                     /organism="unknown"
BASE COUNT     4084 a  11863 c   8356 g   4655 t
ORIGIN

  Query Match            36.7%;  Score 348.2;  DB 6;  Length 28958;
  Best Local Similarity  62.7%;  Pred. No. 2.8e-31;
  Matches  564;  Conservative    0;  Mismatches  323;  Indels   12;  Gaps    1;

Qy          1 GTCTTCGTTTTTCCCGGGCAGGGCCCGCAATGGCCGGGCATGGGAAGGGAACTTCTCGAC 60
              |||||||| ||||| ||||| ||| |||| |||   || ||||         || ||||||
Db       9940 GTCTTCGTCTTTCCTGGGCAAGGCTCGCAGTGGGAAGGGATGGCCCTCTCCCTGCTCGAC 9999

Qy         61 GCTTCCGACGTCTTCCGGGAGAGCGTCCGCGCCTGCGAAGCCGCGTTCGCGCCCTACGTC 120
              | ||    |||||||| |       ||   || |||||   |||| |||| ||  |||||
Db      10000 TCCTCGCCGGTCTTCCGCGCTCAGCTCGAAGCATGCGAGCGCGCGCTCGCTCCTCACGTC 10059

Qy        121 GACTGGTCGGTGGAGCAGGTGTTGCGGGACTCGCCGGACGCTCCCGGGCTGGACCGGGTG 180
              || |||    ||      ||  ||||  |     || ||| |||   || |||||  ||
Db      10060 GAGTGGAGCCTGCTCGCCGTCCTGCGCCGCGACGAGGGCGCCCCCTCCCTCGACCGCGTC 10119

Qy        181 GACGTCGTCCAGCCGACCCTGTTCGCCGTCATGATCTCCCTGGCCGCCCTCTGGCGCTCG 240
              |||||||| |||||  |||| || |||||||| ||||||||||||||||||||||||||||
Db      10120 GACGTCGTACAGCCCGCCCTCTTTGCCGTCATGGTCTCCCTGGCCGCCCTCTGGCGCTCG 10179

Qy        241 CAAGGGGTCGAGCCGTGCGCGGTGCTGGGACACAGCCTGGGCGAGATCGCGGCAGCCCAC 300
              |  || ||||||||   ||| ||  | || |||||||| ||||||||||| || |||  |
Db      10180 CTCGGCGTCGAGCCCGCCGCCGTCGTCGGCCACAGCCAGGGCGAGATCGCCGCCGCCTTC 10239

Qy        301 GTCTCGGGAGGCCTGTCCCTGGCCGACGCCGCACGCGTGGTGACGCTTTGGAGCCAGGCA 360
              ||| | || |  || |||||  |  ||||| || ||| | |   | ||  | ||  | ||
Db      10240 GTCGCAGGCGCTCTCTCCCTCGAGGACGCGGCGCGCATCGCCGCCCTGCGCAGGAAAGCG 10299

Qy        361 CAGACCACCCTTGCCGGGACCGGCGCGCTCGTCTCCGTCGCCGCCACGCCGGATGAGCTC 420
              |  |||||| | | ||| | |||||    | | | ||||||     |    |   || |||
Db      10300 CTCACCACCGTCGGCGGCAACGGCGGCATGGCCGCCGTCGAGCTCGGCGCCTCCGACCTC 10359

Qy        421 CTGCCCCGAATCGCTCCGTGGACCGAGGACAACCCGGCGCGGCTCGCCGTCGCAGCCGTC 480
              | | ||    |||||| |||  |||                ||||| ||  ||| ||||||
Db      10360 CAGACCTACCTCGCTCCCTGGGGCGA------------CAGGCTCTCCACCGCCGCCGTC 10407

Qy        481 AACGGACCCCGGAGCACAGTCGTTTCCGGTGCCCGCGAGGCCGTCGCGGACCTGGTGGCC 540
              ||| | ||| ||   ||  |||| ||||| |  | ||  |||||||  |  ||| | | |
Db      10408 AACAGCCCCAGGGCTACCCTCGTATCCGGCGAGCCCGCCGCCGTCGACGCGCTGCTCGAC 10467

Qy        541 GACCTCACCGCCGCGCAGGTGCGCACGCGCATGATCCCGGTGGACGTTCCCGCCCACTCC 600
              | |||||||||| |  |||||  | | |||| |||||  || |||    || ||||||||
Db      10468 GTCCTCACCGCCACCAAGGTGTTCGCCCGCAAGATCCGCGTCGACTACGCCTCCCACTCC 10527

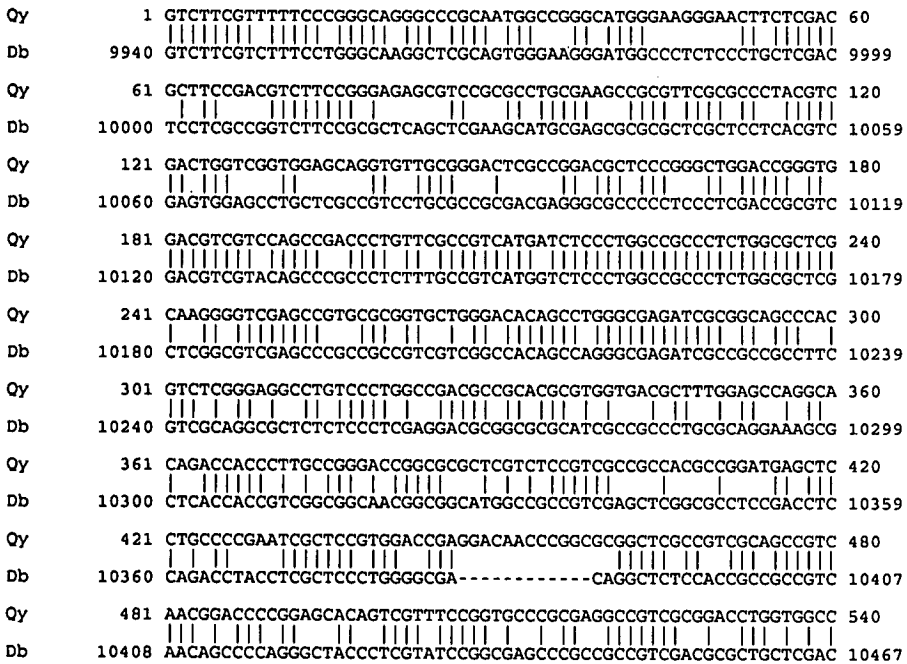
Qy        601 CCCCTGATGTACGCCATCGAGGAACGGGTCGTCAGCGGCCTGCTGCCCATCACCCCACGC 660
              |||  |||| ||||| || | ||  | ||| |   || ||        |||| | || ||
Db      10528 GCCCAGATGGACGCCGTCCAAGACGAGCTCGCCGCAGGTCTAGCCAACATCGCTCCTCGG 10587

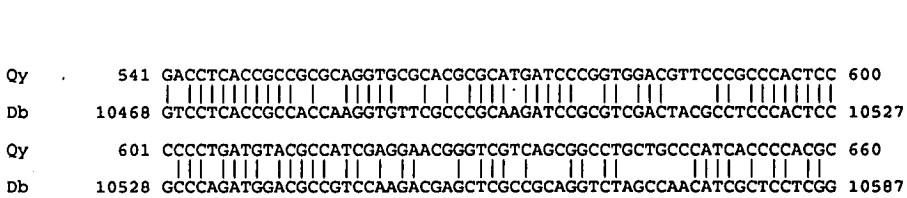
Qy        661 CCCTCCCGCATCCCCTTCCACTCCTCGGTGACCGGCGGCCGCCTCGACACCCGCGAGCTA 720
              | | |   |||  |  | ||  | || ||||||  | | ||||||  |  ||||||
Db      10588 ACGTGCGAGCTCCCTCTTTATTCGACCGTCACCGGCACCAGGCTCGACGGCTCCGAGCTC 10647
```



RESULT 14
I88045
LOCUS I88045 28958 bp DNA linear PAT 10-AUG-1998
DEFINITION Sequence 4 from patent US 5716849.
ACCESSION I88045
VERSION I88045.1 GI:3407985
KEYWORDS .
SOURCE Unknown.
ORGANISM Unknown.
Unclassified.
REFERENCE 1 (bases 1 to 28958)
AUTHORS Ligon,J.M., Schupp,T., Beck,J.J., Hill,D.S., Neff,S. and Ryals,J.A.
TITLE Genes for the biosynthesis of soraphen
JOURNAL Patent: US 5716849-A 4 10-FEB-1998;
FEATURES Location/Qualifiers
source 1. .28958
/organism="unknown"
BASE COUNT 4084 a 11863 c 8356 g 4655 t
ORIGIN

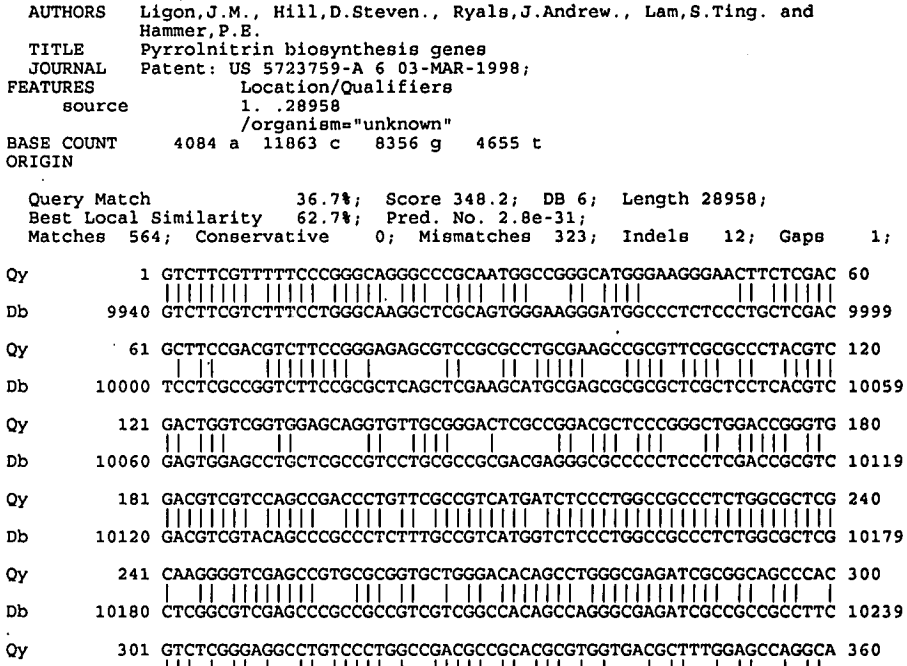
Query Match 36.7%; Score 348.2; DB 6; Length 28958;
Best Local Similarity 62.7%; Pred. No. 2.8e-31;
Matches 564; Conservative 0; Mismatches 323; Indels 12; Gaps 1;





RESULT 15
I90322
LOCUS I90322 28958 bp DNA linear PAT 10-AUG-1998
DEFINITION Sequence 6 from patent US 5723759.
ACCESSION I90322
VERSION I90322.1 GI:3410262
KEYWORDS .
SOURCE Unknown.
ORGANISM Unknown.
Unclassified.
REFERENCE 1 (bases 1 to 28958)
AUTHORS Ligon,J.M., Hill,D.Steven., Ryals,J.Andrew., Lam,S.Ting. and
Hammer,P.E.
TITLE Pyrrolnitrin biosynthesis genes
JOURNAL Patent: US 5723759-A 6 03-MAR-1998;
FEATURES Location/Qualifiers
source 1. .28958
/organism="unknown"
BASE COUNT 4084 a 11863 c 8356 g 4655 t
ORIGIN

Query Match 36.7%; Score 348.2; DB 6; Length 28958;
Best Local Similarity 62.7%; Pred. No. 2.8e-31;
Matches 564; Conservative 0; Mismatches 323; Indels 12; Gaps 1;



```
Qy        361 CAGACCACCCTTGCCGGGACCGGCGCGCTCGTCTCCGTCGCCGCCACGCCGGATGAGCTC 420
              |  ||||| | | ||| | |||||   | | | ||||||    |    |    || |||
Db      10300 CTCACCACCGTCGGCGGCAACGGCGGCATGGCCGCCGTCGAGCTCGGCGCCTCCGACCTC 10359

Qy        421 CTGCCCCGAATCGCTCCGTGGACCGAGGACAACCCGGCGCGGCTCGCCGTCGCAGCCGTC 480
              | | ||    |||||||| |||  |||              |||||  ||  ||| ||||||
Db      10360 CAGACCTACCTCGCTCCCTGGGGCGA----------CAGGCTCTCCACCGCCGCCGTC 10407

Qy        481 AACGGACCCCGGAGCACAGTCGTTTCCGGTGCCCGCGAGGCCGTCGCGGACCTGGTGGCC 540
              ||| | ||| ||   ||  |||| ||||| |  | ||  |||||||  |  ||| | | |
Db      10408 AACAGCCCCAGGGCTACCCTCGTATCCGGCGAGCCCGCCGCCGTCGACGCGCTGCTCGAC 10467

Qy        541 GACCTCACCGCCGCGCAGGTGCGCACGCGCATGATCCCGGTGGACGTTCCCGCCCACTCC 600
              | ||||||||||  |  |||||  | | |||| |||||  || |||    || ||||||||
Db      10468 GTCCTCACCGCCACCAAGGTGTTCGCCCGCAAGATCCGCGTCGACTACGCCTCCCACTCC 10527

Qy        601 CCCCTGATGTACGCCATCGAGGAACGGGTCGTCAGCGGCCTGCTGCCCATCACCCCACGC 660
               ||| |||| |||||  || | ||   | ||| |   || ||      |||| | || ||
Db      10528 GCCCAGATGGACGCCGTCCAAGACGAGCTCGCCGCAGGTCTAGCCAACATCGCTCCTCGG 10587

Qy        661 CCCTCCCGCATCCCCTTCCACTCCTCGGTGACCGGCGGCCGCCTCGACACCCGCGAGCTA 720
              | | |    ||||  |  | ||  | || ||||||  | | ||||||  |  ||||||
Db      10588 ACGTGCGAGCTCCCTCTTTATTCGACCGTCACCGGCACCAGGCTCGACGGCTCCGAGCTC 10647

Qy        721 GACGCGGCGTACTGGTACCGCAACATGTCGAGCACGGTCCGGTTCGAGCCCGCCGCCCGG 780
              ||||  |||||||||| || ||| |   |   || |||| ||||  |  |||  ||  |
Db      10648 GACGGCGCGTACTGGTATCGAAACCTCCGGCAAACCGTCCTGTTCTCGAGCGCGACCGAG 10707

Qy        781 CTGCTTCTGCAGCAGGGGCCCAAGACGTTCGTCGAGATGAGCCCGCACCCGGTGCTGACC 840
              | ||| ||  |  | ||||        | |||||| | |||||  || || |||||  ||
Db      10708 CGGCTCCTCGACGATGGGCATCGCTTCTCCGTCGAGGTCAGCCCCCATCCCGTGCTCACG 10767

Qy        841 ATGGGCCTCCAGGAGCTCGCCCCGGACCTGGGCGACACCACCGGCACCGCCGACACCGT 899
              | | |||||  |||  |  |  |  |         |    ||| |  || || | || |
Db      10768 CTCGCCCTCCGCGAGACCTGCGAGCGCTCACCGCTCGATCCCGTCGTCGTCGGCTCCAT 10826
```

Search completed: June 17, 2003, 17:55:18
Job time : 1849.38 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 17, 2003, 11:26:45 ; Search time 188.583 Seconds
(without alignments)
11320.736 Million cell updates/sec

Title: US-09-914-286-1_COPY_85_1032
Perfect score: 948
Sequence: 1 gtcttcgtttttcccgggca.....gcaccctggaccacttcctg 948

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_101002:*
1: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	948	100.0	11916	22	AAH79279	Streptomyces averm
	2	948	100.0	30690	21	AAA92301	S. avermitilis ave
	3	948	100.0	30690	22	AAH79277	Streptomyces averm
	4	946.4	99.8	12381	21	AAZ58381	Streptomyces averm
	5	354	37.3	65140	22	AAD17184	Streptomyces nours
	6	354	37.3	125401	22	AAD17186	Streptomyces nours
	7	348.2	36.7	28958	18	AAT89956	Sorangium cellulos
	8	348.2	36.7	28958	21	AAA75299	DNA sequence of So
	9	348.2	36.7	49377	19	AAV05287	The soraphen biosy
	10	345	36.4	28598	17	AAT06769	Sorangium cellulos
	11	344	36.3	44377	18	AAT78508	Platenolide syntha
	12	344	36.3	44377	18	AAT80414	Platenolide syntha
	13	337.4	35.6	1010	20	AAV99255	DNA encoding an ac
	14	337.4	35.6	1010	21	AAA72521	DNA sequence of th
	15	319.2	33.7	43280	18	AAT80413	Tylactone synthase
	16	318.4	33.6	29879	14	AAQ46806	eryA region of S.
	17	313.6	33.1	31422	21	AAA92302	S. avermitilis ave
	18	313.6	33.1	31422	22	AAH79278	Streptomyces averm
	19	305	32.2	47981	22	AAF30757	Micromonospora meg
	20	300.6	31.7	7788	22	AAF88335	S. spinosa DNA fra
	21	300.6	31.7	50000	22	AAF88312	S. spinosa DNA fra
	22	300.6	31.7	50000	22	AAF88316	S. spinosa DNA fra
	23	300.6	31.7	80161	20	AAZ21501	DNA fragment of Sa
c	24	293.6	31.0	4403765	22	AAI99683	Mycobacterium tube
c	25	293.6	31.0	4411529	22	AAI99682	Mycobacterium tube
	26	292.4	30.8	27541	22	AAD17185	Streptomyces nours
	27	284	30.0	13842	21	AAZ87297	S. venezuelae macr
	28	284	30.0	36778	21	AAZ87318	S. venezuelae pik
	29	284	30.0	37948	21	AAZ87285	S. venezuelae pik
	30	284	30.0	38506	21	AAA75633	Nucleotide sequenc
	31	284	30.0	38506	21	AAZ56001	Recombinant cosmid
c	32	283.2	29.9	1881	24	AAS18436	Contig 109 DNA enc
	33	279.6	29.5	1035	20	AAV99256	DNA encoding an ac
	34	279.6	29.5	1035	21	AAA72538	DNA encoding the A
	35	275	29.0	16767	22	AAF88339	S. spinosa DNA fra
	36	275	29.0	50000	22	AAF88313	S. spinosa DNA fra
	37	267.2	28.2	11220	21	AAZ87298	S. venezuelae macr
	38	262.6	27.7	4689	21	AAZ87299	S. venezuelae macr
	39	257	27.1	9513	22	AAF88337	S. spinosa DNA fra
	40	256.2	27.0	53789	19	AAV21187	Amycolatopsis medi
	41	253.2	26.7	50937	21	AAA09469	Streptococcus olea
	42	250.6	26.4	4041	21	AAZ87300	S. venezuelae macr
	43	243	25.6	4411529	22	AAI99682	Mycobacterium tube
c	44	233.4	24.6	77536	21	AAA14651	Nucleotide sequenc
	45	233.2	24.6	33529	23	AAS17367	DNA sequence of S.

ALIGNMENTS

RESULT 1
AAH79279
ID AAH79279 standard; DNA; 11916 BP.
XX
AC AAH79279;
XX
DT 04-DEC-2001 (first entry)
XX
DE Streptomyces avermitilis coding sequence derivative SEQ ID NO: 3.
XX
KW Avermectin aglycone synthase; AAS; avermectin derivative;
KW drug production; vetinary drug; pesticide; ds.
XX
OS Synthetic.
XX
FH Key Location/Qualifiers
FT CDS 1..11916
FT /*tag= a
FT /product= "AAG65268"
FT /partial
XX
PN WO200162939-A1.
XX
PD 30-AUG-2001.
XX
PF 23-FEB-2001; 2001WO-JP01381.
XX
PR 24-FEB-2000; 2000JP-0047405.
XX
PA (KYOW) KYOWA HAKKO KOGYO KK.
PA (KITA) KITASATO INST.
XX

PI Endo H, Yamaguchi H, Kanda Y, Hashimoto S, Omura S, Ikeda H;
XX
DR WPI; 2001-582053/65.
DR P-PSDB; AAG65268.
XX
PT New modified avermectin aglycone synthase derived from Streptomyces
PT avermectilis used in production of 22,23-dihydroavermectin B1a used in
PT drugs and pesticides -
XX
PS Claim 10; Page 149-167; 257pp; Japanese.
XX
CC The present invention relates to the production of modified derivatives
CC of avermectin aglycone synthase (AAS) derived from Streptomyces
CC avermitilis. The activity of an acyl carrier protein (ACP),
CC beta-ketoacyl carrier protein synthase (KS), acyltransferase (AT),
CC beta-ketoacyl carrier protein reductase (KR), dehydratase (DH), enoyl
CC reductase (ER) and/or thioesterase (TE) domain may be reduced or
CC suppressed. The process can be used in the production of drugs, vetinary
CC drugs and pesticides. The present sequence is modified version of a
CC fragment of the S. avermitilis genome.
XX
SQ Sequence 11916 BP; 1831 A; 4428 C; 3820 G; 1837 T; 0 other;

Query Match 100.0%; Score 948; DB 22; Length 11916;
Best Local Similarity 100.0%; Pred. No. 2.2e-145;
Matches 948; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 GTCTTCGTTTTTCCCGGGCAGGGCCCGCAATGGCCGGGCATGGGAAGGGAACTTCTCGAC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         85 GTCTTCGTTTTTCCCGGGCAGGGCCCGCAATGGCCGGGCATGGGAAGGGAACTTCTCGAC 144

Qy         61 GCTTCCGACGTCTTCCGGGAGAGCGTCCGCGCCTGCGAAGCCGCGTTCGCGCCCTACGTC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        145 GCTTCCGACGTCTTCCGGGAGAGCGTCCGCGCCTGCGAAGCCGCGTTCGCGCCCTACGTC 204

Qy        121 GACTGGTCGGTGGAGCAGGTGTTGCGGGACTCGCCGGACGCTCCCGGGCTGGACCGGGTG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        205 GACTGGTCGGTGGAGCAGGTGTTGCGGGACTCGCCGGACGCTCCCGGGCTGGACCGGGTG 264

Qy        181 GACGTCGTCCAGCCGACCCTGTTCGCCGTCATGATCTCCCTGGCCGCCCTCTGGCGCTCG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        265 GACGTCGTCCAGCCGACCCTGTTCGCCGTCATGATCTCCCTGGCCGCCCTCTGGCGCTCG 324

Qy        241 CAAGGGGTCGAGCCGTGCGCGGTGCTGGGACACAGCCTGGGCGAGATCGCGGCAGCCCAC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        325 CAAGGGGTCGAGCCGTGCGCGGTGCTGGGACACAGCCTGGGCGAGATCGCGGCAGCCCAC 384

Qy        301 GTCTCGGGAGGCCTGTCCCTGGCCGACGCCGCACGCGTGGTGACGCTTTGGAGCCAGGCA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        385 GTCTCGGGAGGCCTGTCCCTGGCCGACGCCGCACGCGTGGTGACGCTTTGGAGCCAGGCA 444

Qy        361 CAGACCACCCTTGCCGGGACCGGCGCGCTCGTCTCCGTCGCCGCCACGCCGGATGAGCTC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        445 CAGACCACCCTTGCCGGGACCGGCGCGCTCGTCTCCGTCGCCGCCACGCCGGATGAGCTC 504

Qy        421 CTGCCCCGAATCGCTCCGTGGACCGAGGACAACCCGGCGCGGCTCGCCGTCGCAGCCGTC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        505 CTGCCCCGAATCGCTCCGTGGACCGAGGACAACCCGGCGCGGCTCGCCGTCGCAGCCGTC 564

Qy        481 AACGGACCCCGGAGCACAGTCGTTTCCGGTGCCCGCGAGGCCGTCGCGGACCTGGTGGCC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        565 AACGGACCCCGGAGCACAGTCGTTTCCGGTGCCCGCGAGGCCGTCGCGGACCTGGTGGCC 624

Qy        541 GACCTCACCGCCGCGCAGGTGCGCACGCGCATGATCCCGGTGGACGTTCCCGCCCACTCC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        625 GACCTCACCGCCGCGCAGGTGCGCACGCGCATGATCCCGGTGGACGTTCCCGCCCACTCC 684

Qy        601 CCCCTGATGTACGCCATCGAGGAACGGGTCGTCAGCGGCCTGCTGCCCATCACCCCACGC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        685 CCCCTGATGTACGCCATCGAGGAACGGGTCGTCAGCGGCCTGCTGCCCATCACCCCACGC 744

Qy        661 CCCTCCCGCATCCCCTTCCACTCCTCGGTGACCGGCGGCCGCCTCGACACCCGCGAGCTA 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        745 CCCTCCCGCATCCCCTTCCACTCCTCGGTGACCGGCGGCCGCCTCGACACCCGCGAGCTA 804

Qy        721 GACGCGGCGTACTGGTACCGCAACATGTCGAGCACGGTCCGGTTCGAGCCCGCCGCCCGG 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        805 GACGCGGCGTACTGGTACCGCAACATGTCGAGCACGGTCCGGTTCGAGCCCGCCGCCCGG 864

Qy        781 CTGCTTCTGCAGCAGGGGCCCAAGACGTTCGTCGAGATGAGCCCGCACCCGGTGCTGACC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        865 CTGCTTCTGCAGCAGGGGCCCAAGACGTTCGTCGAGATGAGCCCGCACCCGGTGCTGACC 924

Qy        841 ATGGGCCTCCAGGAGCTCGCCCCGGACCTGGGCGACACCACCGGCACCGCCGACACCGTG 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        925 ATGGGCCTCCAGGAGCTCGCCCCGGACCTGGGCGACACCACCGGCACCGCCGACACCGTG 984

Qy        901 ATCATGGGCACGCTGCGCCGCGGCCAGGGCACCCTGGACCACTTCCTG 948
              ||||||||||||||||||||||||||||||||||||||||||||||||
Db        985 ATCATGGGCACGCTGCGCCGCGGCCAGGGCACCCTGGACCACTTCCTG 1032
```

RESULT 2
AAA92301
ID AAA92301 standard; DNA; 30690 BP.
XX
AC AAA92301;
XX
DT 10-JAN-2001 (first entry)
XX
DE S. avermitilis avermectin aglycon synthase DNA aveAI SEQ ID NO:1.
XX
KW Streptomyces avermitilis; avermectin aglycon synthase; biosynthesis;
KW multifunctional enzyme; polyketide; avermectin; veterinary drug;
KW agrochemical; ds.
XX
OS Streptomyces avermitilis.
XX
FH Key Location/Qualifiers
FT CDS 1..11919
FT /*tag= a
FT /note= "avermectin aglycon synthase protein"
FT CDS 11971..30690
FT /*tag= b
FT /note= "avermectin aglycon synthase protein"
XX
PN WO200050605-A1.
XX
PD 31-AUG-2000.
XX
PF 23-FEB-2000; 2000WO-JP01041.
XX
PR 24-FEB-1999; 99JP-0046961.
XX
PA (KITA) KITASATO INST.
XX
PI Omura S, Ikeda H;
XX
DR WPI; 2000-565458/52.
DR P-PSDB; AAB23749, AAB23750.
XX
PT Avermectin aglycone synthase DNA and proteins encoded by all or part of
PT it for the production of avermectin and its derivatives for drug and
PT agrochemical use -
XX
PS Claim 2; Page 66-134; 314pp; Japanese.
XX
CC The present sequence represents DNA which encodes avermectin aglycon
CC synthase proteins. Also described are: (1) polypeptides encoded by all
CC or part of the DNA; (2) expression vectors containing the DNA; (3) host
CC cells transformed by the vectors; (4) preparation of the polypeptides
CC by culture of the transformants; (5) preparation of avermectin aglycon
CC or its derivatives by culture of transformed avermectin-producing
CC microorganisms; and (6) oligonucleotides of 5-60 bases in length
CC containing sense or antisense sequences from the avermectin aglycon
CC synthase DNA. The enzymes are useful for the production of modified

```
CC   forms of avermectin and of the intermediates in its biosynthesis, for
CC   use as drugs, veterinary drugs and agrochemicals.
XX
SQ   Sequence 30690 BP; 5356 A; 12454 C; 8617 G; 4263 T; 0 other;

  Query Match             100.0%;  Score 948;  DB 21;  Length 30690;
  Best Local Similarity   100.0%;  Pred. No. 2e-145;
  Matches  948;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GTCTTCGTTTTTCCCGGGCAGGGCCCGCAATGGCCGGGCATGGGAAGGGAACTTCTCGAC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         85 GTCTTCGTTTTTCCCGGGCAGGGCCCGCAATGGCCGGGCATGGGAAGGGAACTTCTCGAC 144

Qy         61 GCTTCCGACGTCTTCCGGGAGAGCGTCCGCGCCTGCGAAGCCGCGTTCGCGCCCTACGTC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        145 GCTTCCGACGTCTTCCGGGAGAGCGTCCGCGCCTGCGAAGCCGCGTTCGCGCCCTACGTC 204

Qy        121 GACTGGTCGGTGGAGCAGGTGTTGCGGGACTCGCCGGACGCTCCCGGGCTGGACCGGGTG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        205 GACTGGTCGGTGGAGCAGGTGTTGCGGGACTCGCCGGACGCTCCCGGGCTGGACCGGGTG 264

Qy        181 GACGTCGTCCAGCCGACCCTGTTCGCCGTCATGATCTCCCTGGCCGCCCTCTGGCGCTCG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        265 GACGTCGTCCAGCCGACCCTGTTCGCCGTCATGATCTCCCTGGCCGCCCTCTGGCGCTCG 324

Qy        241 CAAGGGGTCGAGCCGTGCGCGGTGCTGGGACACAGCCTGGGCGAGATCGCGGCAGCCCAC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        325 CAAGGGGTCGAGCCGTGCGCGGTGCTGGGACACAGCCTGGGCGAGATCGCGGCAGCCCAC 384

Qy        301 GTCTCGGGAGGCCTGTCCCTGGCCGACGCCGCACGCGTGGTGACGCTTTGGAGCCAGGCA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        385 GTCTCGGGAGGCCTGTCCCTGGCCGACGCCGCACGCGTGGTGACGCTTTGGAGCCAGGCA 444

Qy        361 CAGACCACCCTTGCCGGGACCGGCGCGCTCGTCTCCGTCGCCGCCACGCCGGATGAGCTC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        445 CAGACCACCCTTGCCGGGACCGGCGCGCTCGTCTCCGTCGCCGCCACGCCGGATGAGCTC 504

Qy        421 CTGCCCCGAATCGCTCCGTGGACCGAGGACAACCCGGCGCGGCTCGCCGTCGCAGCCGTC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        505 CTGCCCCGAATCGCTCCGTGGACCGAGGACAACCCGGCGCGGCTCGCCGTCGCAGCCGTC 564

Qy        481 AACGGACCCCGGAGCACAGTCGTTTCCGGTGCCCGCGAGGCCGTCGCGGACCTGGTGGCC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        565 AACGGACCCCGGAGCACAGTCGTTTCCGGTGCCCGCGAGGCCGTCGCGGACCTGGTGGCC 624

Qy        541 GACCTCACCGCCGCGCAGGTGCGCACGCGCATGATCCCGGTGGACGTTCCCGCCCACTCC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        625 GACCTCACCGCCGCGCAGGTGCGCACGCGCATGATCCCGGTGGACGTTCCCGCCCACTCC 684

Qy        601 CCCCTGATGTACGCCATCGAGGAACGGGTCGTCAGCGGCCTGCTGCCCATCACCCCACGC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        685 CCCCTGATGTACGCCATCGAGGAACGGGTCGTCAGCGGCCTGCTGCCCATCACCCCACGC 744

Qy        661 CCCTCCCGCATCCCCTTCCACTCCTCGGTGACCGGCGGCCGCCTCGACACCCGCGAGCTA 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        745 CCCTCCCGCATCCCCTTCCACTCCTCGGTGACCGGCGGCCGCCTCGACACCCGCGAGCTA 804

Qy        721 GACGCGGCGTACTGGTACCGCAACATGTCGAGCACGGTCCGGTTCGAGCCCGCCGCCCGG 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        805 GACGCGGCGTACTGGTACCGCAACATGTCGAGCACGGTCCGGTTCGAGCCCGCCGCCCGG 864

Qy        781 CTGCTTCTGCAGCAGGGGCCCAAGACGTTCGTCGAGATGAGCCCGCACCCGGTGCTGACC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        865 CTGCTTCTGCAGCAGGGGCCCAAGACGTTCGTCGAGATGAGCCCGCACCCGGTGCTGACC 924

Qy        841 ATGGGCCTCCAGGAGCTCGCCCCGGACCTGGGCGACACCACCGGCACCGCCGACACCGTG 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        925 ATGGGCCTCCAGGAGCTCGCCCCGGACCTGGGCGACACCACCGGCACCGCCGACACCGTG 984

Qy        901 ATCATGGGCACGCTGCGCCGCGGCCAGGGCACCCTGGACCACTTCCTG 948
              ||||||||||||||||||||||||||||||||||||||||||||||||
Db        985 ATCATGGGCACGCTGCGCCGCGGCCAGGGCACCCTGGACCACTTCCTG 1032
```

```
RESULT 3
AAH79277
ID   AAH79277 standard; DNA; 30690 BP.
XX
AC   AAH79277;
XX
DT   04-DEC-2001  (first entry)
XX
DE   Streptomyces avermitilis coding sequences SEQ ID NO: 1.
XX
KW   Avermectin aglycone synthase; AAS; avermectin derivative;
KW   drug production; vetinary drug; pesticide; ds.
XX
OS   Streptomyces avermitilis.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..11919
FT                   /*tag= a
FT                   /product= "AAG65264"
FT                   /partial
FT   CDS             11971..30690
FT                   /*tag= b
FT                   /product= "AAG65265"
XX
PN   WO200162939-A1.
XX
PD   30-AUG-2001.
XX
PF   23-FEB-2001; 2001WO-JP01381.
XX
PR   24-FEB-2000; 2000JP-0047405.
XX
PA   (KYOW ) KYOWA HAKKO KOGYO KK.
PA   (KITA ) KITASATO INST.
XX
PI   Endo H,  Yamaguchi H,  Kanda Y,  Hashimoto S,  Omura S,  Ikeda H;
XX
DR   WPI; 2001-582053/65.
DR   P-PSDB; AAG65264, AAG65265.
XX
PT   New modified avermectin aglycone synthase derived from Streptomyces
PT   avermectilis used in production of 22,23-dihydroavermectin B1a used in
PT   drugs and pesticides -
XX
PS   Example 2; Page 58-123; 257pp; Japanese.
XX
CC   The present invention relates to the production of modified derivatives
CC   of avermectin aglycone synthase (AAS) derived from Streptomyces
CC   avermitilis. The activity of an acyl carrier protein (ACP),
CC   beta-ketoacyl carrier protein synthase (KS), acyltransferase (AT),
CC   beta-ketoacyl carrier protein reductase (KR), dehydratase (DH), enoyl
CC   reductase (ER) and/or thioesterase (TE) domain may be reduced or
CC   suppressed. The process can be used in the production of drugs, vetinary
CC   drugs and pesticides. The present sequence is a fragment of the S.
CC   avermitilis genome.
XX
SQ   Sequence 30690 BP; 5355 A; 12455 C; 8617 G; 4263 T; 0 other;

  Query Match             100.0%;  Score 948;  DB 22;  Length 30690;
  Best Local Similarity   100.0%;  Pred. No. 2e-145;
  Matches  948;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GTCTTCGTTTTTCCCGGGCAGGGCCCGCAATGGCCGGGCATGGGAAGGGAACTTCTCGAC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         85 GTCTTCGTTTTTCCCGGGCAGGGCCCGCAATGGCCGGGCATGGGAAGGGAACTTCTCGAC 144

Qy         61 GCTTCCGACGTCTTCCGGGAGAGCGTCCGCGCCTGCGAAGCCGCGTTCGCGCCCTACGTC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        145 GCTTCCGACGTCTTCCGGGAGAGCGTCCGCGCCTGCGAAGCCGCGTTCGCGCCCTACGTC 204

Qy        121 GACTGGTCGGTGGAGCAGGTGTTGCGGGACTCGCCGGACGCTCCCGGGCTGGACCGGGTG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        205 GACTGGTCGGTGGAGCAGGTGTTGCGGGACTCGCCGGACGCTCCCGGGCTGGACCGGGTG 264

Qy        181 GACGTCGTCCAGCCGACCCTGTTCGCCGTCATGATCTCCCTGGCCGCCCTCTGGCGCTCG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        265 GACGTCGTCCAGCCGACCCTGTTCGCCGTCATGATCTCCCTGGCCGCCCTCTGGCGCTCG 324

Qy        241 CAAGGGGTCGAGCCGTGCGCGGTGCTGGGACACAGCCTGGGCGAGATCGCGGCAGCCCAC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        325 CAAGGGGTCGAGCCGTGCGCGGTGCTGGGACACAGCCTGGGCGAGATCGCGGCAGCCCAC 384

Qy        301 GTCTCGGGAGGCCTGTCCCTGGCCGACGCCGCACGCGTGGTGACGCTTTGGAGCCAGGCA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        385 GTCTCGGGAGGCCTGTCCCTGGCCGACGCCGCACGCGTGGTGACGCTTTGGAGCCAGGCA 444

Qy        361 CAGACCACCCTTGCCGGGACCGGCGCGCTCGTCTCCGTCGCCGCCACGCCGGATGAGCTC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        445 CAGACCACCCTTGCCGGGACCGGCGCGCTCGTCTCCGTCGCCGCCACGCCGGATGAGCTC 504

Qy        421 CTGCCCCGAATCGCTCCGTGGACCGAGGACAACCCGGCGCGGCTCGCCGTCGCAGCCGTC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        505 CTGCCCCGAATCGCTCCGTGGACCGAGGACAACCCGGCGCGGCTCGCCGTCGCAGCCGTC 564

Qy        481 AACGGACCCCGGAGCACAGTCGTTTCCGGTGCCCGCGAGGCCGTCGCGGACCTGGTGGCC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        565 AACGGACCCCGGAGCACAGTCGTTTCCGGTGCCCGCGAGGCCGTCGCGGACCTGGTGGCC 624

Qy        541 GACCTCACCGCCGCGCAGGTGCGCACGCGCATGATCCCGGTGGACGTTCCCGCCCACTCC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        625 GACCTCACCGCCGCGCAGGTGCGCACGCGCATGATCCCGGTGGACGTTCCCGCCCACTCC 684

Qy        601 CCCCTGATGTACGCCATCGAGGAACGGGTCGTCAGCGGCCTGCTGCCCATCACCCCACGC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        685 CCCCTGATGTACGCCATCGAGGAACGGGTCGTCAGCGGCCTGCTGCCCATCACCCCACGC 744

Qy        661 CCCTCCCGCATCCCCTTCCACTCCTCGGTGACCGGCGGCCGCCTCGACACCCGCGAGCTA 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        745 CCCTCCCGCATCCCCTTCCACTCCTCGGTGACCGGCGGCCGCCTCGACACCCGCGAGCTA 804

Qy        721 GACGCGGCGTACTGGTACCGCAACATGTCGAGCACGGTCCGGTTCGAGCCCGCCGCCCGG 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        805 GACGCGGCGTACTGGTACCGCAACATGTCGAGCACGGTCCGGTTCGAGCCCGCCGCCCGG 864

Qy        781 CTGCTTCTGCAGCAGGGGCCCAAGACGTTCGTCGAGATGAGCCCGCACCCGGTGCTGACC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        865 CTGCTTCTGCAGCAGGGGCCCAAGACGTTCGTCGAGATGAGCCCGCACCCGGTGCTGACC 924

Qy        841 ATGGGCCTCCAGGAGCTCGCCCCGGACCTGGGCGACACCACCGGCACCGCCGACACCGTG 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        925 ATGGGCCTCCAGGAGCTCGCCCCGGACCTGGGCGACACCACCGGCACCGCCGACACCGTG 984

Qy        901 ATCATGGGCACGCTGCGCCGCGGCCAGGGCACCCTGGACCACTTCCTG 948
              ||||||||||||||||||||||||||||||||||||||||||||||||
Db        985 ATCATGGGCACGCTGCGCCGCGGCCAGGGCACCCTGGACCACTTCCTG 1032
```

RESULT 4
AAZ58381
ID AAZ58381 standard; DNA; 12381 BP.
XX
AC AAZ58381;
XX
DT 23-MAY-2000 (first entry)
XX
DE Streptomyces avermitilis avermectin polyketide synthase modules 1+2.
XX
KW Polyketide synthase; avermectin; insecticide; ss.
XX
OS Streptomyces avermitilis.
XX
PN WO200001827-A2.
XX
PD 13-JAN-2000.
XX
PF 06-JUL-1999; 99WO-GB02158.
XX
PR 06-JUL-1998; 98GB-0014622.
XX
PA (BIOT-) BIOTICA TECHNOLOGY LTD.
PA (PFIZ) PFIZER INC.
XX
PI Kellenberger JL, Leadlay PF, Staunton J, Stutzman-Engwall KJ;
PI McArthur HAI;
XX
DR WPI; 2000-182117/16.
XX
PT Mutated Type I polyketide synthase containing a polylinker site in an
PT extension module for replacement of a reductive loop sequence, for
PT producing polyketides, e.g. B1 avermectin -
XX
PS Disclosure; Fig 7a-f; 75pp; English.
XX
CC The present sequence is that of DNA encoding the first 2 modules
CC of the avermectin polyketide synthase (PKS) of Streptomyces
CC avermitilis. The invention relates to nucleic acids encoding a
CC Type I PKS such as avermectin in which a polylinker with multiple
CC restriction sites replaces or 1 more PKS genes encoding enzymes
CC associated with reduction. Novel PKS are provided in which in
CC which the reductive loop in a selected module of the Type I PKS is
CC replaced with the equivalent segment from the same or different
CC PKS gene cluster or by a mutated or synthetic segment. Vectors and
CC host cells, and methods for producing novel polyketides by
CC culturing host cells are claimed. The polyketides obtained are
CC useful as antibiotics and insecticides. Fermentation products
CC containing C22-C23 dihydroavermectin, ivermectin and B1
CC avermectins are claimed.
XX
SQ Sequence 12381 BP; 1884 A; 4561 C; 4005 G; 1931 T; 0 other;

Query Match 99.8%; Score 946.4; DB 21; Length 12381;
Best Local Similarity 99.9%; Pred. No. 3.9e-145;
Matches 947; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy          1 GTCTTCGTTTTTCCCGGGCAGGGCCCGCAATGGCCGGGCATGGGAAGGGAACTTCTCGAC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        902 GTCTTCGTTTTTCCCGGGCAGGGCCCGCAATGGCCGGGCATGGGAAGGGAACTTCTCGAC 961

Qy         61 GCTTCCGACGTCTTCCGGGAGAGCGTCCGCGCCTGCGAAGCCGCGTTCGCGCCCTACGTC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        962 GCTTCCGACGTCTTCCGGGAGAGCGTCCGCGCCTGCGAAGCCGCGTTCGCGCCCTACGTC 1021

Qy        121 GACTGGTCGGTGGAGCAGGTGTTGCGGGACTCGCCGGACGCTCCCGGGCTGGACCGGGTG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1022 GACTGGTCGGTGGAGCAGGTGTTGCGGGACTCGCCGGACGCTCCCGGGCTGGACCGGGTG 1081

Qy        181 GACGTCGTCCAGCCGACCCTGTTCGCCGTCATGATCTCCCTGGCCGCCCTCTGGCGCTCG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1082 GACGTCGTCCAGCCGACCCTGTTCGCCGTCATGATCTCCCTGGCCGCCCTCTGGCGCTCG 1141

Qy        241 CAAGGGGTCGAGCCGTGCGCGGTGCTGGGACACAGCCTGGGCGAGATCGCGGCAGCCCAC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1142 CAAGGGGTCGAGCCGTGCGCGGTGCTGGGACACAGCCTGGGCGAGATCGCGGCAGCCCAC 1201

Qy        301 GTCTCGGGAGGCCTGTCCCTGGCCGACGCCGCACGCGTGGTGACGCTTTGGAGCCAGGCA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1202 GTCTCGGGAGGCCTGTCCCTGGCCGACGCCGCACGCGTGGTGACGCTTTGGAGCCAGGCA 1261

Qy        361 CAGACCACCCTTGCCGGGACCGGCGCGCTCGTCTCCGTCGCCGCCACGCCGGATGAGCTC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1262 CAGACCACCCTTGCCGGGACCGGCGCGCTCGTCTCCGTCGCCGCCACGCCGGATGAGCTC 1321

Qy        421 CTGCCCCGAATCGCTCCGTGGACCGAGGACAACCCGGCGCGGCTCGCCGTCGCAGCCGTC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1322 CTGCCCCGAATCGCTCCGTGGACCGAGGACAACCCGGCGCGGCTCGCCGTCGCAGCCGTC 1381

Qy        481 AACGGACCCCGGAGCACAGTCGTTTCCGGTGCCCGCGAGGCCGTCGCGGACCTGGTGGCC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db       1382 AACGGACCCCCGGAGCACAGTCGTTTCCGGTGCCCGCGAGGCCGTCGCGGACCTGGTGGCC 1441

Qy        541 GACCTCACCGCCGCGCGCAGGTGCGCACGCGCATGATCCCGGTGGACGTTCCCGCCCACTCC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1442 GACCTCACCGCCGCGCGCAGGTGCGCACGCGCATGATCCCGGTGGACGTTCCCGCCCACTCC 1501

Qy        601 CCCCTGATGTACGCCATCGAGGAACGGGTCGTCAGCGGCCTGCTGCCCATCACCCCACGC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1502 CCCCTGATGTACGCCATCGAGGAACGGGTCGTCAGCGGCCTGCTGCCCATCACCCCACGC 1561

Qy        661 CCCTCCCGCATCCCCTTCCACTCCTCGGTGACCGGCGGCCGCCTCGACACCCGCGAGCTA 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1562 CCCTCCCGCATCCCCTTCCACTCCTCGGTGACCGGCGGCCGCCTCGACACCCGCGAGCTA 1621

Qy        721 GACGCGGCGTACTGGTACCGCAACATGTCGAGCACGGTCCGGTTCGAGCCCGCCGCCCGG 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1622 GACGCGGCGTACTGGTACCGCAACATGTCGAGCACGGTCCGGTTCGAGCCCGCCGCCCGG 1681

Qy        781 CTGCTTCTGCAGCAGGGGCCCAAGACGTTCGTCGAGATGAGCCCGCACCCGGTGCTGACC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1682 CTGCTTCTGCAGCAGGGGCCCAAGACGTTCGTCGAGATGAGCCCGCACCCGGTGCTGACC 1741

Qy        841 ATGGGCCTCCAGGAGCTCGCCCCGGACCTGGGCGACACCACCGGCACCGCCGACACCGTG 900
              |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db       1742 ATGGGCCTCCAGGAGCTCGCCGCGGACCTGGGCGACACCACCGGCACCGCCGACACCGTG 1801

Qy        901 ATCATGGGCACGCTGCGCCGCGGCCAGGGCACCCTGGACCACTTCCTG 948
              ||||||||||||||||||||||||||||||||||||||||||||||||
Db       1802 ATCATGGGCACGCTGCGCCGCGGCCAGGGCACCCTGGACCACTTCCTG 1849


RESULT 5
AAD17184
ID   AAD17184 standard; DNA; 65140 BP.
XX
AC   AAD17184;
XX
DT   29-NOV-2001  (first entry)
XX
DE   Streptomyces noursei nys1 DNA of nystatin PKS gene cluster.
XX
KW   Polyketide synthase; PKS; macrolide; nystatin; PKS gene cluster;
KW   antifungal; antibiotic; nys1; ds.
XX
OS   Streptomyces noursei.
XX
FH   Key             Location/Qualifiers
FT   CDS             complement (1..1035)
FT                   /*tag= a
FT                   /product= "NysD2 partial protein"
FT                   /note= "CDS does not include stop codon"
FT   CDS             complement (1056..2576)
FT                   /*tag= b
FT                   /product= "NysD1 protein"
FT   CDS             2806..6906
FT                   /*tag= c
FT                   /product= "NysA protein"
FT   CDS             6952..16530
FT                   /*tag= d
FT                   /product= "NysB protein"
FT   CDS             16550..49840
FT                   /*tag= e
FT                   /product= "NysC protein"
FT   CDS             50260..51015
FT                   /*tag= f
FT                   /product= "NysE protein"
FT   CDS             51405..54305
FT                   /*tag= g
FT                   /product= "NysR1 protein"
FT   CDS             54329..57190
FT                   /*tag= h
FT                   /product= "NysR2 protein"
FT                   /note= "CDS does not include start codon"
FT   CDS             57180..59963
FT                   /*tag= i
FT                   /product= "NysR3 protein"
FT   CDS             60415..61047
FT                   /*tag= j
FT                   /product= "NysR4 (short) protein"
FT                   /note= "CDS does not include start codon"
FT   CDS             61736..62497
FT                   /*tag= k
FT                   /product= "NysR5 protein"
FT                   /note= "CDS does not include start codon"
FT   CDS             complement (62551..63615)
FT                   /*tag= l
FT                   /product= "ORF2 protein"
FT                   /note= "CDS does not include start codon"
FT   CDS             63765..64961
FT                   /*tag= m
FT                   /product= "ORF1 protein"
XX
PN   WO200159126-A2.
XX
PD   16-AUG-2001.
XX
PF   08-FEB-2001; 2001WO-GB00509.
XX
PR   08-FEB-2000; 2000GB-0002840.
PR   10-APR-2000; 2000GB-0008786.
PR   14-APR-2000; 2000GB-0009387.
XX
PA   (UYNO-) UNIV NORGES TEKNISK NATURVITENSKAPELIGE.
PA   (SNTF ) SINTEF STIFTELSEN IND TEK FORSK.
PA   (ALPH-) ALPHARMA AS.
PA   (SINV-) SINVENT AS.
PA   (DZIE/) DZIEGLEWSKA H.
PA   (ZOTC/) ZOTCHEV S B.
PA   (SEKU/) SEKUROVA O N.
PA   (FJAE/) FJAERVIK E.
PA   (BRAU/) BRAUTASET T.
PA   (STRO/) STROM A R.
XX
PI   Zotchev SB,  Sekurova ON,  Fjaervik E,  Brautaset T,  Strom AR;
PI   Valla S,  Ellingsen TE,  Sletta H,  Gulliksen O;
XX
DR   WPI; 2001-557614/62.
DR   P-PSDB; AAE10125, AAE10126, AAE10127, AAE10128, AAE10129, AAE10130,
DR   AAE10131, AAE10132, AAE10133, AAE10134, AAE10135, AAE10136, AAE10137.
XX
PT   New nystatin polyketide synthase polynucleotides and polypeptides,
PT   useful as antibiotics and antifungals -
XX
PS   Claim 2; Page 116-151; 266pp; English.
XX
CC   The present invention relates to the cloning and sequencing of the gene
CC   cluster encoding a modular type I polyketide synthase (PKS) enzyme
CC   involved in the biosynthesis of the macrolide antibiotic nystatin.
CC   The nystatin PKS is useful as antifungal antibiotics. The present
CC   sequence is a Streptomyces noursei nys1 DNA of nystatin PKS gene cluster.
XX
SQ   Sequence 65140 BP; 8270 A; 25171 C; 22273 G; 9426 T; 0 other;

  Query Match             37.3%;  Score 354;  DB 22;  Length 65140;
  Best Local Similarity   62.6%;  Pred. No. 2.6e-49;
  Matches  593;  Conservative    0;  Mismatches  340;  Indels   15;  Gaps    2;

Qy          1 GTCTTCGTTTTTCCCGGGCAGGGCCCGCAATGGCCGGGCATGGGAAGGGAACTTCTCGAC 60
              |||||||| || ||||| |||||  | || |||  ||| |||||  | |||||  |||| |
Db      13291 GTCTTCGTCTTCCCCGGCCAGGGATCCCAGTGGTGGGGAATGGGCCGCGAACTCCTCGCC 13350

Qy         61 GCTTCCGACGTCTTCCGGGAGAGCGTCCGCGCCTGCGAAGCCGCGTTCGCGCCCTACGTC 120
               | ||||| || |||||          ||    | |||||    | ||  |||| || ||||||
Db      13351 ACGTCCGAGGTGTTCCGCACCGCGATCGATGACTGCGCGACGGCCCTCGCCCCGTACGTC 13410

Qy        121 GACTGGTCGGTGGAGCAGGTGTTGCGGGACTCGCCGGACGCTCCCGGGCTGGACCGGGTG 180
```

```
Db      13411 GACTGGTCGCTGCACGACGTCCTGGCCGGCGAGGGCGACCCCGCCCTGCTGGAGCGGGTG 13470

Qy        181 GACGTCGTCCAGCCGACCCTGTTCGCCGTCATGATCTCCCTGGCCGCCCTCTGGCGCTCG 240

Db      13471 GACGTGGTCCAGCCCGCGCTGTTCGCCATGATGGTCGGGCTGTCCGCGCTCTGGCGCTCC 13530

Qy        241 CAAGGGGTCGAGCCGTGCGCGGTGCTGGGACACAGCCTGGGCGAGATCGCGGCAGCCCAC 300

Db      13531 CACGGCGTCGTCCCGGCGGCCGTGGTCGGCCACTCGCAGGGCGAGATCGCCGCGGCCTGC 13590

Qy        301 GTCTCGGGAGGCCTGTCCCTGGCCGACGCCGCACGCGTGGTGACGCTTTGGAGCCAGGCA 360

Db      13591 GTCGCCGGAGCCCTCAGCCTGGCCGACGCCGCCCGCGTGGTGGCGCTGCGCAGCCAGGCA 13650

Qy        361 CAGACCACCCTTGCCGGGACCGGCGCGCTCGTCTCCGTCGCCGCCACGCCGGATGAGCTC 420

Db      13651 CTGCCGCAACTGTCCGGACGCGGCGGCATGATGTCGGTCTCCGCCCCCGTAGAGCGGGTC 13710

Qy        421 CTGCCCCGAATCGCTCCGTGGACCGAGGACAACCCGGCGCGGCTCGCCGTCGCAGCCGTC 480

Db      13711 ACCGCACTCCTCGCCCCGTGG----------CAGGAGGCGCTGTCCGTCGCCGCGGTC 13758

Qy        481 AACGGACCCCGGAGCACAGTCGTTTCCGGTGCCCGCGAGGCCGTCGCGGACCTGGTGGCC 540

Db      13759 AACGGCCCCTCGTCCGTGGTCGTCTCCGGCGACACCGACGCGCTCGACGCCCTGCACACC 13818

Qy        541 GACCTCACCGCCGCGCAGGTGCGCACGCGCATGATCCCGGTGGACGTTCCCGCCCACTCC 600

Db      13819 GCCTGCCAGGAACAGGGCGTGCGGGCCCGCAAGGTGTCCGTGGACTACGCCTCGCACGGG 13878

Qy        601 CCCCTGATGTACGCCATCGAGGAACGGGTCGTCAGCGGCCTGCTGCCCATCACCCCACGC 660

Db      13879 CGGCACGTCGAGGCCGTCCGCGACGAACTCGCCCGCGTCCTCGCGCCGGTCGACCCGCGC 13938

Qy        661 CCCTCCCGCATCCCCTTCCACTCCTCGGTGACCGGCGGCCGCCTCGACACCCGCGAGCTA 720

Db      13939 GCCCCCGAGGTGCCGTTCTACTCGACGGTCACCGGCGACCGCGTGGACGACGCCGCCTTC 13998

Qy        721 GACGCGGCGTACTGGTACCGCAACATGTCGAGCACGGTCCGGTTCGAGCCCGCCGCCCGG 780

Db      13999 GACGGCGCCTACTGGTACACCAACCTCCGCCAGACCGTCCGCATGGAGGAGGCCACCCGC 14058

Qy        781 CTGCTTCTGCAGCAGGGGCCCAAGACGTTCGTCGAGATGAGCCCGCACCCGGTGCTGACC 840

Db      14059 GCCCTCCTCGCCGCCGGACACCGCGTCTTCATCGAGGTCAGCCCGCACCCGGTGCTCGCC 14118

Qy        841 ATGGGCCTCCAGGAGCTCGCCCCGGACCTGGGCGACACCACCGGCACCGCCGACACCGTG 900

Db      14119 GCCCCGATCCAGGAGACGCAGGAGGCCGTAGCGGAGGCCACCGGCGGGTCCG---CGGTG 14175

Qy        901 ATCATGGGCACGCTGCGCCGCGGCCAGGGCACCCTGGACCACTTCCTG 948

Db      14176 GTCCTCGGCTCGCTCCGCCGCGACGAGGGCGGCCCGCGGCGCTTCCTG 14223


RESULT 6
AAD17186
ID   AAD17186 standard; DNA; 125401 BP.
XX
AC   AAD17186;
XX
DT   29-NOV-2001  (first entry)
XX
DE   Streptomyces noursei nystatin PKS gene cluster DNA.
XX
KW   Polyketide synthase; PKS; macrolide; nystatin; PKS gene cluster;
KW   antifungal; antibiotic; ds.
XX
OS   Streptomyces noursei.
XX
FH   Key             Location/Qualifiers
FT   CDS             6337..34771
FT                   /*tag= a
FT                   /product= "NysI complete protein"
FT   CDS             34792..51099
FT                   /*tag= b
FT                   /product= "NysJ protein"
FT   CDS             51155..57355
FT                   /*tag= c
FT                   /product= "NysK protein"
FT   CDS             57503..58687
FT                   /*tag= d
FT                   /product= "NysL protein"
FT   CDS             complement (58786..58980)
FT                   /*tag= e
FT                   /product= "NysM protein"
FT                   /note= "CDS does not include start codon"
FT   CDS             complement (59045..60241)
FT                   /*tag= f
FT                   /product= "NysN protein"
FT                   /note= "CDS does not include start codon"
FT   CDS             complement (60238..61296)
FT                   /*tag= g
FT                   /product= "NysD2 complete protein"
FT   CDS             120628..121308
FT                   /*tag= h
FT                   /product= "NysR4 (long) protein"
XX
PN   WO200159126-A2.
XX
PD   16-AUG-2001.
XX
PF   08-FEB-2001; 2001WO-GB00509.
XX
PR   08-FEB-2000; 2000GB-0002840.
PR   10-APR-2000; 2000GB-0008786.
PR   14-APR-2000; 2000GB-0009387.
XX
PA   (UYNO-) UNIV NORGES TEKNISK NATURVITENSKAPELIGE.
PA   (SNTF ) SINTEF STIFTELSEN IND TEK FORSK.
PA   (ALPH-) ALPHARMA AS.
PA   (SINV-) SINVENT AS.
PA   (DZIE/) DZIEGLEWSKA H.
PA   (ZOTC/) ZOTCHEV S B.
PA   (SEKU/) SEKUROVA O N.
PA   (FJAE/) FJAERVIK E.
PA   (BRAU/) BRAUTASET T.
PA   (STRO/) STROM A R.
XX
PI   Zotchev SB,  Sekurova ON,  Fjaervik E,  Brautaset T,  Strom AR;
PI   Valla S,  Ellingsen TE,  Sletta H,  Gulliksen O;
XX
DR   WPI; 2001-557614/62.
DR   P-PSDB; AAE10143, AAE10144, AAE10145, AAE10146, AAE10147, AAE10148,
DR   AAE10149, AAE10150.
XX
PT   New nystatin polyketide synthase polynucleotides and polypeptides,
PT   useful as antibiotics and antifungals -
XX
PS   Claim 1; Page 188-254; 266pp; English.
XX
CC   The present invention relates to the cloning and sequencing of the gene
CC   cluster encoding a modular type I polyketide synthase (PKS) enzyme
CC   involved in the biosynthesis of the macrolide antibiotic nystatin.
CC   The nystatin PKS is useful as antifungal antibiotics. The present
CC   sequence is a Streptomyces noursei nystatin PKS gene cluster DNA.
XX
SQ   Sequence 125401 BP; 15664 A; 49692 C; 42871 G; 17174 T; 0 other;

  Query Match             37.3%;  Score 354;  DB 22;  Length 125401;
  Best Local Similarity   62.6%;  Pred. No. 2.4e-49;
  Matches  593;  Conservative    0;  Mismatches  340;  Indels   15;  Gaps    2;

Qy          1 GTCTTCGTTTTTCCCGGGCAGGGCCCGCAATGGCCGGGCATGGGAAGGGAACTTCTCGAC 60
```

```
Db      73552 GTCTTCGTCTTCCCCGGCCAGGGATCCCAGTGGTGGGGAATGGGCCGCGAACTCCTCGCC 73611

Qy         61 GCTTCCGACGTCTTCCGGGAGAGCGTCCGCGCCTGCGAAGCCGCGTTCGCGCCCTACGTC 120

Db      73612 ACGTCCGAGGTGTTCCGCACCGCGATCGATGACTGCGCGACGGCCCTCGCCCCGTACGTC 73671

Qy        121 GACTGGTCGGTGGAGCAGGTGTTGCGGGACTCGCCGGACGCTCCCGGGCTGGACCGGGTG 180

Db      73672 GACTGGTCGCTGCACGACGTCCTGGCCGGCGAGGGCGACCCCGCCCTGCTGGAGCGGGTG 73731

Qy        181 GACGTCGTCCAGCCGACCCTGTTCGCCGTCATGATCTCCCTGGCCGCCCTCTGGCGCTCG 240

Db      73732 GACGTGGTCCAGCCCGCGCTGTTCGCCATGATGGTCGGGCTGTCCGCGCTCTGGCGCTCC 73791

Qy        241 CAAGGGGTCGAGCCGTGCGCGGTGCTGGGACACAGCCTGGGCGAGATCGCGGCAGCCCAC 300

Db      73792 CACGGCGTCGTCCCGGCGGCCGTGGTCGGCCACTCGCAGGGCGAGATCGCCGCGGCCTGC 73851

Qy        301 GTCTCGGGAGGCCTGTCCCTGGCCGACGCCGCACGCGTGGTGACGCTTTGGAGCCAGGCA 360

Db      73852 GTCGCCGGAGCCCTCAGCCTGGCCGACGCCGCCCGCGTGGTGGCGCTGCGCAGCCAGGCA 73911

Qy        361 CAGACCACCCTTGCCGGGACCGGCGCGCTCGTCTCCGTCGCCGCCACGCCGGATGAGCTC 420

Db      73912 CTGCCGCAACTGTCCGGACGCGGCGGCATGATGTCGGTCTCCGCCCCCGTAGAGCGGGTC 73971

Qy        421 CTGCCCCGAATCGCTCCGTGGACCGAGGACAACCCGGCGCGGCTCGCCGTCGCAGCCGTC 480

Db      73972 ACCGCACTCCTCGCCCCGTGG----------CAGGAGGCGCTGTCCGTCGCCGCGGTC 74019

Qy        481 AACGGACCCCGGAGCACAGTCGTTTCCGGTGCCCGCGAGGCCGTCGCGGACCTGGTGGCC 540

Db      74020 AACGGCCCCTCGTCCGTGGTCGTCTCCGGCGACACCGACGCGCTCGACGCCCTGCACACC 74079

Qy        541 GACCTCACCGCCGCGCAGGTGCGCACGCGCATGATCCCGGTGGACGTTCCCGCCCACTCC 600

Db      74080 GCCTGCCAGGAACAGGGCGTGCGGGCCCGCAAGGTGTCCGTGGACTACGCCTCGCACGGG 74139

Qy        601 CCCCTGATGTACGCCATCGAGGAACGGGTCGTCAGCGGCCTGCTGCCCATCACCCCACGC 660

Db      74140 CGGCACGTCGAGGCCGTCCGCGACGAACTCGCCCGCGTCCTCGCGCCGGTCGACCCGCGC 74199

Qy        661 CCCTCCCGCATCCCCTTCCACTCCTCGGTGACCGGCGGCCGCCTCGACACCCGCGAGCTA 720

Db      74200 GCCCCCGAGGTGCCGTTCTACTCGACGGTCACCGGCGACCGCGTGGACGACGCCGCCTTC 74259

Qy        721 GACGCGGCGTACTGGTACCGCAACATGTCGAGCACGGTCCGGTTCGAGCCCGCCGCCCGG 780

Db      74260 GACGGCGCCTACTGGTACACCAACCTCCGCCAGACCGTCCGCATGGAGGAGGCCACCCGC 74319

Qy        781 CTGCTTCTGCAGCAGGGGCCCAAGACGTTCGTCGAGATGAGCCCGCACCCGGTGCTGACC 840

Db      74320 GCCCTCCTCGCCGCCGGACACCGCGTCTTCATCGAGGTCAGCCCGCACCCGGTGCTCGCC 74379

Qy        841 ATGGGCCTCCAGGAGCTCGCCCCGGACCTGGGCGACACCACCGGCACCGCCGACACCGTG 900

Db      74380 GCCCCGATCCAGGAGACGCAGGAGGCCGTAGCGGAGGCCACCGGCGGGTCCG---CGGTG 74436

Qy        901 ATCATGGGCACGCTGCGCCGCGGCCAGGGCACCCTGGACCACTTCCTG 948

Db      74437 GTCCTCGGCTCGCTCCGCCGCGACGAGGGCGGCCCGCGGCGCTTCCTG 74484


RESULT 7
AAT89956
ID   AAT89956 standard; DNA; 28958 BP.
XX
AC   AAT89956;
XX
DT   12-MAR-1998  (first entry)
XX
DE   Sorangium cellulosum soraphen gene cluster genomic DNA.
XX
KW   Soraphen; polyketide synthase; PKS; antipathogenic; antibiotic;
KW   transgenic plant; phytopathogen; resistance; ss.
XX
OS   Sorangium cellulosum.
XX
PN   US5662898-A.
XX
PD   02-SEP-1997.
XX
PF   01-JUN-1995;   95US-0457342.
XX
PR   20-AUG-1990;   90US-0570184.
PR   02-JUL-1992;   92US-0908284.
PR   31-AUG-1992;   92US-0937648.
PR   01-JUL-1993;   93US-0087636.
PR   08-JUN-1994;   94US-0258261.
XX
PA   (CIBA ) CIBA GEIGY CORP.
XX
PI   Beck JJ,  Gaffney TD,  Hammer PE,  Hill DS,  Lam ST;
PI   Ligon JM,  Ryals JA,  Schupp T,  Uknes SJ;
XX
DR   WPI; 1997-447901/41.
XX
PT   Protecting plants against pathogens with genetically transformed
PT   biological control agent - which expresses all polypeptide(s)
PT   involved in pyrrolnitrin biosynthetic pathway
XX
PS   Ex 16; Column 104-128; 88pp; English.
XX
CC   This genomic DNA sequence encodes the Soraphen gene cluster. This
CC   sequence encodes two open reading frames (ORF's), the positions of which
CC   are not given in the specification. ORF1 is approximately 25.5 kb in
CC   size and encodes 5 biosynthetic modules homologous to the erythromycin
CC   genes of Saccharopolyspora erythraea. Each module contains a
CC   beta-ketoacylsynthase (KS), an acyltransferase (AT), a ketoreductase
CC   (KR) and an acyl carrier protein (ACP) domain. ORF2 is immediately
CC   adjacent to ORF1 and is thought to contain 3 modules. Soraphen is a
CC   type I polyketide synthase (PKS) which has antibiotic activity.
CC   Transgenic plants containing such antipathogenic genes like those
CC   encoded in the Soraphen cluster should have enhanced resistance to
CC   attack by phytopathogens.
XX
SQ   Sequence 28958 BP; 4084 A; 11863 C; 8356 G; 4655 T; 0 other;

  Query Match           36.7%;  Score 348.2;  DB 18;  Length 28958;
  Best Local Similarity   62.7%;  Pred. No. 2.5e-48;
  Matches  564;  Conservative    0;  Mismatches  323;  Indels   12;  Gaps    1;

Qy          1 GTCTTCGTTTTTCCCGGGCAGGGCCCGCAATGGCCGGGCATGGGAAGGGAACTTCTCGAC 60

Db       9940 GTCTTCGTCTTTCCTGGGCAAGGCTCGCAGTGGGAAGGGATGGCCCTCTCCCTGCTCGAC 9999

Qy         61 GCTTCCGACGTCTTCCGGGAGAGCGTCCGCGCCTGCGAAGCCGCGTTCGCGCCCTACGTC 120

Db      10000 TCCTCGCCGGTCTTCCGCGCTCAGCTCGAAGCATGCGAGCGCGCGCTCGCTCCTCACGTC 10059

Qy        121 GACTGGTCGGTGGAGCAGGTGTTGCGGGACTCGCCGGACGCTCCCGGGCTGGACCGGGTG 180

Db      10060 GAGTGGAGCCTGCTCGCCGTCCTGCGCCGCGACGAGGGCGCCCCCTCCCTCGACCGCGTC 10119

Qy        181 GACGTCGTCCAGCCGACCCTGTTCGCCGTCATGATCTCCCTGGCCGCCCTCTGGCGCTCG 240

Db      10120 GACGTCGTACAGCCCGCCCTCTTTGCCGTCATGGTCTCCCTGGCCGCCCTCTGGCGCTCG 10179

Qy        241 CAAGGGGTCGAGCCGTGCGCGGTGCTGGGACACAGCCTGGGCGAGATCGCGGCAGCCCAC 300

Db      10180 CTCGGCGTCGAGCCCGCCGCCGTCGTCGGCCACAGCCAGGGCGAGATCGCCGCCGCCTTC 10239

Qy        301 GTCTCGGGAGGCCTGTCCCTGGCCGACGCCGCACGCGTGGTGACGCTTTGGAGCCAGGCA 360

Db      10240 GTCGCAGGCGCTCTCTCCCTCGAGGACGCGGCGCGCATCGCCGCCCTGCGCAGGAAAGCG 10299
```

```
Qy        361 CAGACCACCCTTGCCGGGACCGGCGCGCTCGTCTCCGTCGCCGCCACGCCGGATGAGCTC 420
              |  ||||||  | |  |||  | |||||   | | |  ||||||     |     |    || |||
Db      10300 CTCACCACCGTCGGCGGCAACGGCGGCATGGCCGCCGTCGAGCTCGGCGCCTCCGACCTC 10359

Qy        421 CTGCCCCGAATCGCTCCGTGGACCGAGGACAACCCGGCGCGGCTCGCCGTCGCAGCCGTC 480
              | | ||    |||||| |||  |||                |||||  ||  ||| ||||||
Db      10360 CAGACCTACCTCGCTCCCTGGGGCGA------------CAGGCTCTCCACCGCCGCCGTC 10407

Qy        481 AACGGACCCCGGAGCACAGTCGTTTCCGGTGCCCGCGAGGCCGTCGCGGACCTGGTGGCC 540
              ||| | ||| ||   ||  |||| |||||  |  | ||  |||||||  |  ||| | | |
Db      10408 AACAGCCCCAGGGCTACCCTCGTATCCGGCGAGCCCGCCGCCGTCGACGCGCTGCTCGAC 10467

Qy        541 GACCTCACCGCCGCGCAGGTGCGCACGCGCATGATCCCGGTGGACGTTCCCGCCCACTCC 600
              | ||||||||||  |  |||||  | | |||| |||||  ||  |||     || ||||||||
Db      10468 GTCCTCACCGCCACCAAGGTGTTCGCCCGCAAGATCCGCGTCGACTACGCCTCCCACTCC 10527

Qy        601 CCCCTGATGTACGCCATCGAGGAACGGGTCGTCAGCGGCCTGCTGCCCATCACCCCACGC 660
               ||| |||| ||||| ||  | ||   | ||| |   ||  ||       |||| | ||  ||
Db      10528 GCCCAGATGGACGCCGTCCAAGACGAGCTCGCCGCAGGTCTAGCCAACATCGCTCCTCGG 10587

Qy        661 CCCTCCCGCATCCCCTTCCACTCCTCGGTGACCGGCGGCCGCCTCGACACCCGCGAGCTA 720
              | | |    ||||  |  | ||  | ||  ||||||  | |  ||||||  |  ||||||
Db      10588 ACGTGCGAGCTCCCTCTTTATTCGACCGTCACCGGCACCAGGCTCGACGGCTCCGAGCTC 10647

Qy        721 GACGCGGCGTACTGGTACCGCAACATGTCGAGCACGGTCCGGTTCGAGCCCGCCGCCCGG 780
              ||||  ||||||||||| || ||| |    |   || |||| ||||  |  ||| ||  |
Db      10648 GACGGCGCGTACTGGTATCGAAACCTCCGGCAAACCGTCCTGTTCTCGAGCGCGACCGAG 10707

Qy        781 CTGCTTCTGCAGCAGGGGCCCAAGACGTTCGTCGAGATGAGCCCGCACCCGGTGCTGACC 840
              | ||| ||  |  | ||||        | |||||| | ||||| || || ||||| ||
Db      10708 CGGCTCCTCGACGATGGGCATCGCTTCTCCGTCGAGGTCAGCCCCCATCCCGTGCTCACG 10767

Qy        841 ATGGGCCTCCAGGAGCTCGCCCCGGACCTGGGCGACACCACCGGCACCGCCGACACCGT 899
              | | |||||  |||  |  |  | |         |   ||| |  || || | || |
Db      10768 CTCGCCCTCCGCGAGACCTGCGAGCGCTCACCGCTCGATCCCGTCGTCGTCGGCTCCAT 10826
```

RESULT 8
AAA75299
ID AAA75299 standard; DNA; 28958 BP.
XX
AC AAA75299;
XX
DT 15-JAN-2001 (first entry)
XX
DE DNA sequence of Sorangium cellulosum soraphen gene region.
XX
KW Pyrrolnitrin; PrnA; PrnB; PrnC; PrnD; antibiotic; biocontrol;
KW fungal respiratory electron transport inhibitor; transgenic plant;
KW antipathogenic substance; biosynthetic gene; fungal resistance;
KW phytopathogen; soraphen; ss.
XX
OS Sorangium cellulosum.
XX
PN US6117670-A.
XX
PD 12-SEP-2000.
XX
PF 24-FEB-1998; 98US-0028934.
XX
PR 30-MAY-1995; 95WO-IB00414.
PR 08-JUN-1994; 94US-0258261.
PR 09-OCT-1996; 96US-0729214.
XX
PA (NOVS) NOVARTIS FINANCE CORP.
XX
PI Ligon JM, Lam ST, Hammer PE, Van Pee K, Kirner S, Young TR;
PI Hill DS;
XX
DR WPI; 2000-593649/56.
XX
PT Pyrrolnitrin biosynthesis genes and enzymes, useful for producing
PT pyrrolnitrin for increasing plant resistance to phytopathogens, e.g.
PT fungi or nematodes, and for developing microbial hosts useful as
PT biocontrol agents -
XX
PS Example 14; Column 121-146; 124pp; English.
XX
CC The present sequence represents a Sorangium cellulosum soraphen gene
CC cluster. The specification describes pyrrolnitrin gene regions. This
CC region encodes enzymes required in the biosynthetic pathway of
CC pyrrolnitrin. The PrnA enzyme catalyses the conversion of tryptophan
CC to 7-chlorotryptophan. The PrnB enzyme catalyses the conversion of
CC 7-chlorotryptophan to monodechloroaminopyrrolnitrin. PrnC enzyme
CC catalyses the conversion of monodechloroaminopyrrolnitrin to
CC aminopyrrolnitrin. PrnD enzyme catalyses the conversion of
CC aminopyrrolnitrin to pyrrolnitrin. Pyrrolnitrin has antibiotic activity,
CC and inhibits a broad range of fungi. It is a fungal respiratory electron
CC transport inhibitor. The genes are useful for developing a host as a
CC biocontrol agent, for protecting plants against a pathogen and for
CC producing large, uniform amounts of pyrrolnitrin. The genes are also
CC useful for producing transgenic plants that express antipathogenic
CC substance (APS) biosynthetic genes. The genes and enzymes are useful
CC for increasing the resistance of host plants to disease caused by
CC phytopathogens, e.g. fungi, bacteria or nematodes. The microbial hosts
CC are useful for controlling or inhibiting the growth of a phytopathogen.
XX
SQ Sequence 28958 BP; 4084 A; 11863 C; 8356 G; 4655 T; 0 other;

Query Match 36.7%; Score 348.2; DB 21; Length 28958;
Best Local Similarity 62.7%; Pred. No. 2.5e-48;
Matches 564; Conservative 0; Mismatches 323; Indels 12; Gaps 1;

```
Qy          1 GTCTTCGTTTTTCCCGGGCAGGGCCCGCAATGGCCGGGCATGGGAAGGGAACTTCTCGAC 60
              |||||||| ||||| ||||| ||| |||| |||    || ||||         || ||||||
Db       9940 GTCTTCGTCTTTCCTGGGCAAGGCTCGCAGTGGGAAGGGATGGCCCTCTCCCTGCTCGAC 9999

Qy         61 GCTTCCGACGTCTTCCGGGAGAGCGTCCGCGCCTGCGAAGCCGCGTTCGCGCCCTACGTC 120
              | ||    |||||||| |    ||  || |||||   |||| |||| ||  |||||
Db      10000 TCCTCGCCGGTCTTCCGCGCTCAGCTCGAAGCATGCGAGCGCGCGCTCGCTCCTCACGTC 10059

Qy        121 GACTGGTCGGTGGAGCAGGTGTTGCGGGACTCGCCGGACGCTCCCGGGCTGGACCGGGTG 180
              || |||    ||      ||  ||||   |     || ||| |||   || ||||| ||
Db      10060 GAGTGGAGCCTGCTCGCCGTCCTGCGCCGCGACGAGGGCGCCCCCTCCCTCGACCGCGTC 10119

Qy        181 GACGTCGTCCAGCCGACCCTGTTCGCCGTCATGATCTCCCTGGCCGCCCTCTGGCGCTCG 240
              |||||||| |||||  |||| || ||||||||| ||||||||||||||||||||||||||
Db      10120 GACGTCGTACAGCCCGCCCTCTTTGCCGTCATGGTCTCCCTGGCCGCCCTCTGGCGCTCG 10179

Qy        241 CAAGGGGTCGAGCCGTGCGCGGTGCTGGGACACAGCCTGGGCGAGATCGCGGCAGCCCAC 300
              |  || ||||||||   ||| ||  | || ||||||| |||||||||||| || |||  |
Db      10180 CTCGGCGTCGAGCCCGCCGCCGTCGTCGGCCACAGCCAGGGCGAGATCGCCGCCGCCTTC 10239

Qy        301 GTCTCGGGAGGCCTGTCCCTGGCCGACGCCGCACGCGTGGTGACGCTTTGGAGCCAGGCA 360
              ||| | || |  || ||||| |  ||||| || ||| | |   | ||  | ||  | ||
Db      10240 GTCGCAGGCGCTCTCTCCCTCGAGGACGCGGCGCGCATCGCCGCCCTGCGCAGGAAAGCG 10299

Qy        361 CAGACCACCCTTGCCGGGACCGGCGCGCTCGTCTCCGTCGCCGCCACGCCGGATGAGCTC 420
              |  ||||||  | |  |||  | |||||   | | |  ||||||     |     |    || |||
Db      10300 CTCACCACCGTCGGCGGCAACGGCGGCATGGCCGCCGTCGAGCTCGGCGCCTCCGACCTC 10359

Qy        421 CTGCCCCGAATCGCTCCGTGGACCGAGGACAACCCGGCGCGGCTCGCCGTCGCAGCCGTC 480
              | | ||    |||||| |||  |||                |||||  ||  ||| ||||||
Db      10360 CAGACCTACCTCGCTCCCTGGGGCGA------------CAGGCTCTCCACCGCCGCCGTC 10407

Qy        481 AACGGACCCCGGAGCACAGTCGTTTCCGGTGCCCGCGAGGCCGTCGCGGACCTGGTGGCC 540
              ||| | ||| ||   ||  |||| |||||  |  | ||  |||||||  |  ||| | | |
Db      10408 AACAGCCCCAGGGCTACCCTCGTATCCGGCGAGCCCGCCGCCGTCGACGCGCTGCTCGAC 10467

Qy        541 GACCTCACCGCCGCGCAGGTGCGCACGCGCATGATCCCGGTGGACGTTCCCGCCCACTCC 600
              | ||||||||||  |  |||||  | | |||| |||||  ||  |||     || ||||||||
Db      10468 GTCCTCACCGCCACCAAGGTGTTCGCCCGCAAGATCCGCGTCGACTACGCCTCCCACTCC 10527

Qy        601 CCCCTGATGTACGCCATCGAGGAACGGGTCGTCAGCGGCCTGCTGCCCATCACCCCACGC 660
               ||| |||| ||||| ||  | ||   | ||| |   ||  ||       |||| | ||  ||
Db      10528 GCCCAGATGGACGCCGTCCAAGACGAGCTCGCCGCAGGTCTAGCCAACATCGCTCCTCGG 10587
```

```
Qy        661 CCCTCCCGCATCCCCTTCCACTCCTCGGTGACCGGCGGCCGCCTCGACACCCGCGAGCTA 720
              | | |    |||| |  | ||  | || ||||||  | | ||||||  |  ||||||
Db      10588 ACGTGCGAGCTCCCTCTTTATTCGACCGTCACCGGCACCAGGCTCGACGGCTCCGAGCTC 10647

Qy        721 GACGCGGCGTACTGGTACCGCAACATGTCGAGCACGGTCCGGTTCGAGCCCGCCGCCCGG 780
              ||||  |||||||||||| || ||| |   |   || |||| |||| |  |||  ||  |
Db      10648 GACGGCGCGTACTGGTATCGAAACCTCCGGCAAACCGTCCTGTTCTCGAGCGCGACCGAG 10707

Qy        781 CTGCTTCTGCAGCAGGGGCCCAAGACGTTCGTCGAGATGAGCCCGCACCCGGTGCTGACC 840
              | ||| ||  |  | ||||       | ||||||| | ||||| || || ||||| ||
Db      10708 CGGCTCCTCGACGATGGGCATCGCTTCTCCGTCGAGGTCAGCCCCCATCCCGTGCTCACG 10767

Qy        841 ATGGGCCTCCAGGAGCTCGCCCCGGACCTGGGCGACACCACCGGCACCGCCGACACCGT 899
              | | |||||  |||  |  |  |  |         |   ||| |  || || | || |
Db      10768 CTCGCCCTCCGCGAGACCTGCGAGCGCTCACCGCTCGATCCCGTCGTCGTCGGCTCCAT 10826
```

RESULT 9
AAV05287
ID AAV05287 standard; DNA; 49377 BP.
XX
AC AAV05287;
XX
DT 21-MAY-1998 (first entry)
XX
DE The soraphen biosynthesis gene cluster from Sorangium cellulosum.
XX
KW Polyketide synthase; PKS; biosynthesis; soraphen; SorR; SorA; SorB;
KW SorM; biosynthetic module; beta-ketoacylsynthase; acyltransferase;
KW ketoreductase; beta-ketone processing domain; cytostatic agent;
KW antimicrobial agent; phytopathogenic fungi; transgenic plant;
KW biological control; ss.
XX
OS Sorangium cellulosum.
XX

FH	Key	Location/Qualifiers
FT	CDS	383..760
FT		/*tag= a
FT		/product= SorR
FT		/note= "gene product highly homologous to the reductase domains of type I PKSs such as eryA from Saccharopolyspora erythraea"
FT	CDS	927..19874
FT		/*tag= b
FT		/product= SorA
FT		/note= "gene product is highly homologous to type I PKSs that are known to be involved in the synthesis of polyketide compounds"
FT	misc_feature	942..7115
FT		/*tag= c
FT		/note= "module 1 of SorA"
FT	misc_feature	7203..12884
FT		/*tag= d
FT		/note= "module 2 of SorA"
FT	misc_feature	13455..19616
FT		/*tag= e
FT		/note= "module 3 of SorA"
FT	misc_feature	19871..46318
FT		/*tag= f
FT	CDS	19871..46318
FT		/*tag= g
FT		/product= SorB
FT		/note= "gene product is highly homologous to type I PKS genes"
FT	misc_feature	19870..24556
FT		/*tag= h
FT		/note= "module 1 of SorB"
FT	misc_feature	24638..30820
FT		/*tag= i
FT		/note= "module 2 of SorB"
FT	misc_feature	30881..35446
FT		/*tag= j
FT		/note= "module 3 of SorB"
FT	misc_feature	35528..40114
FT		/*tag= k
FT		/note= "module 4 of SorB"
FT	misc_feature	40190..46318
FT		/*tag= l
FT		/note= "module 5 of SorB"
FT	CDS	46851..47891
FT		/*tag= m
FT		/product= SorM
FT		/note= "gene product is homologous to the methyltransferase from Streptomyces hygroscopicus that is involved in the synthesis of the polyketide rappamicin"

XX
PN US5716849-A.
XX
PD 10-FEB-1998.
XX
PF 14-DEC-1996; 96US-0764233.
XX
PR 24-AUG-1993; 93WO-US07954.
PR 08-JUN-1994; 94US-0258261.
PR 09-OCT-1996; 96US-0729214.
XX
PA (NOVS) NOVARTIS FINANCE CORP.
XX
PI Beck JJ, Hill DS, Ligon JM, Neff S, Ryals JA, Schupp T;
XX
DR WPI; 1998-158369/14.
XX
PT DNA encoding Sorangium cellulosum polypeptide(s) - used for, e.g.
PT biosynthesis of soraphen useful as antimicrobial agent against
PT phytopathogenic fungi.
XX
PS Claim 2; Columns 47-90; 64pp; English.
XX
CC The present sequence contains a cluster of genes that encode polyketide
CC synthases (PKSs) that are involved in the synthesis of soraphens in
CC Sorangium cellulosum. The proteins encoded by the present sequence are
CC SorR, SorA, SorB and SorM. SorA and SorB contain biosynthetic modules
CC which contain a beta-ketoacylsynthase, an acyltransferase, a
CC ketoreductase and an acyl carrier protein domain, as well as beta-ketone
CC processing domains. S. cellulosum soraphens are useful as a cytostatic
CC and antimicrobial agent active against phytopathogenic fungi.
CC Soraphen-producing transgenic plants or biological control agents can
CC also be produced, which may reduce crop losses and nutritional
CC deprivation for local populations in many parts of the world.
XX
SQ Sequence 49377 BP; 7247 A; 19522 C; 14477 G; 8131 T; 0 other;

Query Match 36.7%; Score 348.2; DB 19; Length 49377;
Best Local Similarity 62.7%; Pred. No. 2.3e-48;
Matches 564; Conservative 0; Mismatches 323; Indels 12; Gaps 1;

```
Qy          1 GTCTTCGTTTTTCCCGGGCAGGGCCCGCAATGGCCGGGCATGGGAAGGGAACTTCTCGAC 60
              |||||||| ||||| ||||| ||| |||| |||   || ||||        || ||||||
Db      26243 GTCTTCGTCTTTCCTGGGCAAGGCTCGCAGTGGGAAGGGATGGCCCTCTCCCTGCTCGAC 26302

Qy         61 GCTTCCGACGTCTTCCGGGAGAGCGTCCGCGCCTGCGAAGCCGCGTTCGCGCCCTACGTC 120
               | ||   |||||||| |      ||   || |||||   |||| |||| ||  |||||
Db      26303 TCCTCGCCGGTCTTCCGCGCTCAGCTCGAAGCATGCGAGCGCGCGCTCGCTCCTCACGTC 26362

Qy        121 GACTGGTCGGTGGAGCAGGTGTTGCGGGACTCGCCGGACGCTCCCGGGCTGGACCGGGTG 180
              || |||    ||       ||  ||||   |      || ||| |||   || ||||| ||
Db      26363 GAGTGGAGCCTGCTCGCCGTCCTGCGCCGCGACGAGGGCGCCCCCTCCCTCGACCGCGTC 26422

Qy        181 GACGTCGTCCAGCCGACCCTGTTCGCCGTCATGATCTCCCTGGCCGCCCTCTGGCGCTCG 240
              |||||||| |||||  |||| || ||||||||| ||||||||||||||||||||||||||
Db      26423 GACGTCGTACAGCCCGCCCTCTTTGCCGTCATGGTCTCCCTGGCCGCCCTCTGGCGCTCG 26482

Qy        241 CAAGGGGTCGAGCCGTGCGCGGTGCTGGGACACAGCCTGGGCGAGATCGCGGCAGCCCAC 300
              |  || ||||||||   ||| ||  | || ||||||| |||||||||||| || |||  |
```

```
Db      26483 CTCGGCGTCGAGCCCGCCGCCGTCGTCGGCCACAGCCAGGGCGAGATCGCCGCCGCCTTC 26542

Qy        301 GTCTCGGGAGGCCTGTCCCTGGCCGACGCCGCACGCGTGGTGACGCTTTGGAGCCAGGCA 360

Db      26543 GTCGCAGGCGCTCTCTCCCTCGAGGACGCGGCGCGCATCGCCGCCCTGCGCAGGAAAGCG 26602

Qy        361 CAGACCACCCTTGCCGGGACCGGCGCGCTCGTCTCCGTCGCCGCCACGCCGGATGAGCTC 420

Db      26603 CTCACCACCGTCGGCGGCAACGGCGGCATGGCCGCCGTCGAGCTCGGCGCCTCCGACCTC 26662

Qy        421 CTGCCCCGAATCGCTCCGTGGACCGAGGACAACCCGGCGCGGCTCGCCGTCGCAGCCGTC 480

Db      26663 CAGACCTACCTCGCTCCCTGGGGCGA----------CAGGCTCTCCACCGCCGCCGTC 26710

Qy        481 AACGGACCCCGGAGCACAGTCGTTTCCGGTGCCCGCGAGGCCGTCGCGGACCTGGTGGCC 540

Db      26711 AACAGCCCCAGGGCTACCCTCGTATCCGGCGAGCCCGCCGCCGTCGACGCGCTGCTCGAC 26770

Qy        541 GACCTCACCGCCGCGCAGGTGCGCACGCGCATGATCCCGGTGGACGTTCCCGCCCACTCC 600

Db      26771 GTCCTCACCGCCACCAAGGTGTTCGCCCGCAAGATCCGCGTCGACTACGCCTCCCACTCC 26830

Qy        601 CCCCTGATGTACGCCATCGAGGAACGGGTCGTCAGCGGCCTGCTGCCCATCACCCCACGC 660

Db      26831 GCCCAGATGGACGCCGTCCAAGACGAGCTCGCCGCAGGTCTAGCCAACATCGCTCCTCGG 26890

Qy        661 CCCTCCCGCATCCCCTTCCACTCCTCGGTGACCGGCGGCCGCCTCGACACCCGCGAGCTA 720

Db      26891 ACGTGCGAGCTCCCTCTTTATTCGACCGTCACCGGCACCAGGCTCGACGGCTCCGAGCTC 26950

Qy        721 GACGCGGCGTACTGGTACCGCAACATGTCGAGCACGGTCCGGTTCGAGCCCGCCGCCCGG 780

Db      26951 GACGGCGCGTACTGGTATCGAAACCTCCGGCAAACCGTCCTGTTCTCGAGCGCGACCGAG 27010

Qy        781 CTGCTTCTGCAGCAGGGGCCCAAGACGTTCGTCGAGATGAGCCCGCACCCGGTGCTGACC 840

Db      27011 CGGCTCCTCGACGATGGGCATCGCTTCTCCGTCGAGGTCAGCCCCCATCCCGTGCTCACG 27070

Qy        841 ATGGGCCTCCAGGAGCTCGCCCCGGACCTGGGCGACACCACCGGCACCGCCGACACCGT 899

Db      27071 CTCGCCCTCCGCGAGACCTGCGAGCGCTCACCGCTCGATCCCGTCGTCGTCGGCTCCAT 27129


RESULT 10
AAT06769
ID   AAT06769 standard; DNA; 28598 BP.
XX
AC   AAT06769;
XX
DT   15-OCT-1996  (first entry)
XX
DE   Sorangium cellulosum soraphen gene cluster.
XX
KW   Antipathogenic substance; soraphen; phenylpyrrole; antibiotic;
KW   fungicide; pesticide; myxobacterium; ss.
XX
OS   Sorangium cellulosum.
XX
PN   WO9533818-A2.
XX
PD   14-DEC-1995.
XX
PF   30-MAY-1995;   95WO-IB00414.
XX
PR   08-JUN-1994;   94US-0258261.
XX
PA   (CIBA ) CIBA GEIGY AG.
XX
PI   Beck JJ,  Gaffney TD,  Hammer PE,  Hill DS,  Lam ST;
PI   Ligon J,  Ryals JA, Schupp T,  Uknes SJ;
XX
DR   WPI; 1996-040226/04.
XX
PT   New genes for biosynthesis of anti-pathogenic substances - pref.
PT   pyrrolnitrin and soraphen, useful for disease control in plants
XX
PS   Claim 3; Page 140-158; 190pp; English.
XX
CC   This is the soraphen gene cluster from S. cellulosum.  The gene
CC   cluster may be expressed recombinantly to produce soraphen, or
CC   expressed in a transgenic plant for disease-resistance.
XX
SQ   Sequence 28598 BP; 4031 A; 11703 C; 8263 G; 4601 T; 0 other;

  Query Match             36.4%;  Score 345;  DB 17;  Length 28598;
  Best Local Similarity   62.5%;  Pred. No. 8.2e-48;
  Matches  562;  Conservative    0;  Mismatches  325;  Indels   12;  Gaps    1;

Qy          1 GTCTTCGTTTTTCCCGGGCAGGGCCCGCAATGGCCGGGCATGGGAAGGGAACTTCTCGAC 60

Db      25192 GTCTTCGTCTTTCCTGGGCAAGGCTCGCAGTGGGAAGGGATGGCCCTCTCGCTGCTCGAC 25251

Qy         61 GCTTCCGACGTCTTCCGGGAGAGCGTCCGCGCCTGCGAAGCCGCGTTCGCGCCCTACGTC 120

Db      25252 TCCTCGCCCGTCTTCCGCGCTCAGCTCGAAGCATGCGAGCGCGCGCTCGCTCCTCACGTC 25311

Qy        121 GACTGGTCGGTGGAGCAGGTGTTGCGGGACTCGCCGGACGCTCCCGGGCTGGACCGGGTG 180

Db      25312 GAGTGGAGCCTGCTCGCCGTCCTGCGCCGCGACGAGGGCGCCCCCTCCCTCGACCGCGTC 25371

Qy        181 GACGTCGTCCAGCCGACCCTGTTCGCCGTCATGATCTCCCTGGCCGCCCTCTGGCGCTCG 240

Db      25372 GACGTCGTACAGCCCGCCCTCTTTGCCGTCATGGTCTCCCTGGCGGCCCTCTGGCGCTCG 25431

Qy        241 CAAGGGGTCGAGCCGTGCGCGGTGCTGGGACACAGCCTGGGCGAGATCGCGGCAGCCCAC 300

Db      25432 CTCGGCGTAGAGCCCGCCGCCGTCGTCGGCCACAGTCAGGGCGAGATCGCCGCCGCCTTC 25491

Qy        301 GTCTCGGGAGGCCTGTCCCTGGCCGACGCCGCACGCGTGGTGACGCTTTGGAGCCAGGCA 360

Db      25492 GTCGCAGGCGCTCTCTCCCTCGAGGACGCGGCCCGCATCGCCGCCCTGCGCAGCAAAGCG 25551

Qy        361 CAGACCACCCTTGCCGGGACCGGCGCGCTCGTCTCCGTCGCCGCCACGCCGGATGAGCTC 420

Db      25552 CTCACCACCGTCGCCGGCAACGGGGCCATGGCCGCCGTCGAGCTCGGCGCCTCCGACCTC 25611

Qy        421 CTGCCCCGAATCGCTCCGTGGACCGAGGACAACCCGGCGCGGCTCGCCGTCGCAGCCGTC 480

Db      25612 CAGACCTACCTCGCTCCCTGGGGCGA----------CAGGCTCTCCATCGCCGCCGTC 25659

Qy        481 AACGGACCCCGGAGCACAGTCGTTTCCGGTGCCCGCGAGGCCGTCGCGGACCTGGTGGCC 540

Db      25660 AACAGCCCCAGGGCCACGCTCGTGTCCGGCGAGCCCGCCGCCATCGACGCGCTGATCGAC 25719

Qy        541 GACCTCACCGCCGCGCAGGTGCGCACGCGCATGATCCCGGTGGACGTTCCCGCCCACTCC 600

Db      25720 TCGCTCACCGCAGCGCAGGTCTTCGCCCGAAAAGTCCGCGTCGACTACGCCTCCCACTCC 25779

Qy        601 CCCCTGATGTACGCCATCGAGGAACGGGTCGTCAGCGGCCTGCTGCCCATCACCCCACGC 660

Db      25780 GCCCAGATGGACGCCGTCCAAGACGAGCTCGCCGCAGGTCTAGCCAACATCGCTCCTCGG 25839

Qy        661 CCCTCCCGCATCCCCTTCCACTCCTCGGTGACCGGCGGCCGCCTCGACACCCGCGAGCTA 720

Db      25840 ACGTGCGAGCTCCCTCTTTATTCGACCGTCACCGGCACCAGGCTCGACGGCTCCGAGCTC 25899

Qy        721 GACGCGGCGTACTGGTACCGCAACATGTCGAGCACGGTCCGGTTCGAGCCCGCCGCCCGG 780

Db      25900 GACGGCGCGTACTGGTATCGAAACCTCCGGCAAACCGTCCTGTTCTCGAGCGCGACCGAG 25959

Qy        781 CTGCTTCTGCAGCAGGGGCCCAAGACGTTCGTCGAGATGAGCCCGCACCCGGTGCTGACC 840

Db      25960 CGGCTCCTCGACGATGGGCATCGCTTCTTCGTCGAGGTCAGCCCCCATCCCGTGCTCACG 26019

Qy        841 ATGGGCCTCCAGGAGCTCGCCCCGGACCTGGGCGACACCACCGGCACCGCCGACACCGT 899
```

```
Db      26020 CTCGCCCTCCGCGAGACCTGCGAGCGCTCACCGCTCGATCCCGTCGTCGTCGGCTCCAT 26078

RESULT 11
AAT78508
ID   AAT78508 standard; DNA; 44377 BP.
XX
AC   AAT78508;
XX
DT   26-FEB-1998  (first entry)
XX
DE   Platenolide synthase gene cluster.
XX
KW   Platenolide synthase gene cluster; platenolide production; srmG gene;
KW   multi-functional protein; macrolide antibiotic; spiramycin; ss.
XX
OS   Streptomyces ambofaciens.
XX
FH   Key             Location/Qualifiers
FT   CDS             350..14002
FT                   /*tag= a
FT                   /transl_except= (pos:350..352, aa:Met)
FT                   /note= "ORF1 encodes protein shown in AAW23716"
FT   CDS             14046..20036
FT                   /*tag= b
FT                   /note= "ORF2 encodes protein shown in AAW23717"
FT   CDS             20110..31284
FT                   /*tag= c
FT                   /transl_except= (pos:20111..20113, aa:Met)
FT                   /note= "ORF3 encodes protein shown in AAW23718"
FT   CDS             31329..36071
FT                   /*tag= d
FT                   /note= "ORF4 encodes protein shown in AAW23719"
FT   CDS             36155..41830
FT                   /*tag= e
FT                   /note= "ORF5 encodes protein shown in AAW23720"
XX
PN   EP791656-A2.
XX
PD   27-AUG-1997.
XX
PF   19-FEB-1997;   97EP-0301066.
XX
PR   22-FEB-1996;   96US-0012050.
XX
PA   (ELIL ) LILLY & CO ELI.
XX
PI   Burgett SG,  Kuhstoss SA,  Rao RN,  Richardson MA;
PI   Rosteck PR;
XX
DR   WPI; 1997-418047/39.
DR   P-PSDB; AAW23716-W23720.
XX
PT   DNA encoding Streptomyces ambofaciens platenolide synthase domain -
PT   for production of spiramycin-related polyketide antibiotics
XX
PS   Claim 9; Pages 8-33; 81pp; English.
XX
CC   This sequence represents the platenolide synthase gene cluster of the
CC   invention. This sequence is referred to as the srmG gene, and was
CC   isolated from Streptomyces ambofaciens. This sequence encodes the
CC   multi-functional proteins which direct the synthesis of the polyketide
CC   platenolide. Platenolide is the basic building block of the macrolide
CC   antibiotic spiramycin. The DNA can be used to produce compounds
CC   exhibiting antibiotic activity based on the platenolide structure,
CC   including specifically the macrolide antibiotic spiramycin and spiramycin
CC   analogues and derivatives. Modifications of the platenolide synthase DNA
CC   sequence can be made so as to change the number and type of carboxylic
CC   acids incorporated into the growing polyketide chain and to change the
CC   kind of post-condensation processing that is conducted.
XX
SQ   Sequence 44377 BP; 4965 A; 15552 C; 17381 G; 6479 T; 0 other;

  Query Match           36.3%;  Score 344;  DB 18;  Length 44377;
  Best Local Similarity   62.5%;  Pred. No. 1.1e-47;
  Matches  599;  Conservative    0;  Mismatches  335;  Indels   24;  Gaps    3;

Qy          1 GTCTTCGTTTTTCCCGGGCAGGGCCCGCAATGGCCGGGCATGGGAAGGGAACTTCT---- 56
              || ||||| || ||||| |||||| |||| ||||| |  ||||      |  || ||
Db      26374 GTGTTCGTCTTCCCCGGCCAGGGCTCGCAGTGGCCCGAGATGGCCGACGGGCTGCTGGCC 26433

Qy         57 -----CGACGCTTCCGACGTCTTCCGGGAGAGCGTCCGCGCCTGCGAAGCCGCGTTCGCG 111
                  |  ||  |||| |  ||||| ||||| || ||||||||||||    |||| ||  |
Db      26434 CGCTCCAGCGGCTCCGGCTCCTTCCTGGAGACCGCCCGCGCCTGCGACCTCGCGCTCCGG 26493

Qy        112 CCCTACGTCGACTGGTCGGTGGAGCAGGTGTTGCGGGACTCGCCGGACGCTCCCGGGCTG 171
              ||| || ||| |||||| ||   | | ||  ||||        || | ||| |||  |||
Db      26494 CCCCACCTCGGCTGGTCCGTCCTGGACGTACTGCGCCGGGAACCCGGCGCGCCCTCGCTC 26553

Qy        172 GACCGGGTGGACGTCGTCCAGCCGACCCTGTTCGCCGTCATGATCTCCCTGGCCGCCCTC 231
              |||||||| |||||  || |||||     |||||| || | ||| |||| || ||||
Db      26554 GACCGGGTCGACGTGGTGCAGCCCGTGCTGTTCACCATGATGGTCTCGCTCGCCGAGACG 26613

Qy        232 TGGCGCTCGCAAGGGGTCGAGCCGTGCGCGGTGCTGGGACACAGCCTGGGCGAGATCGCG 291
              ||||| ||||  || ||||| |||  ||||||  | ||  |||  || ||||||||||||
Db      26614 TGGCGTTCGCTGGGCGTCGAACCGGCCGCGGTCGTCGGTCACTCCCAGGGCGAGATCGCC 26673

Qy        292 GCAGCCCACGTCTCGGGAGGCCTGTCCCTGGCCGACGCCGCACGCGTGGTGACGCTTTGG 351
              || ||| ||||| | || | |||| | |||| |||||| || ||| | ||  | || |
Db      26674 GCCGCCTACGTCGCCGGCGCCCTGACGCTGGACGACGCGGCGCGCATCGTCGCCCTGCGC 26733

Qy        352 AGCCAGGCACAGACCACCCTTGCCGGGACCGGCGCGCTCGTCTCCGTCGCCGCCACGCCG 411
              ||||||||   |       || |||||  |  ||||   | ||| ||||  ||     |
Db      26734 AGCCAGGCGTGGCTGCGGCTGGCCGGCAAGGGCGGCATGGTCGCCGTGACCCTGTCCGAA 26793

Qy        412 GATGAGCTCCTGCCCCGAATCGCTCCGTGGACCGAGGACAACCCGGCGCGGCTCGCCGTC 471
              || || |  |||||  | |  || |||| |||             ||||||||||||
Db      26794 CGCGACCTGCGTCCCCGCCTGGAGCCCTGGAGCGA------------CCGGCTCGCCGTC 26841

Qy        472 GCAGCCGTCAACGGACCCCGGAGCACAGTCGTTTCCGGTGCCCGCGAGGCCGTCGCGGAC 531
              || |||||||||||| |||  || |   | ||| ||||| | ||  || ||| | |||||
Db      26842 GCCGCCGTCAACGGCCCCGAGACCTGCGCCGTCTCCGGGGACCCGGACGCCCTGGCGGAG 26901

Qy        532 CTGGTGGCCGACCTCACCGCCGCGCAGGTGCGCACGCGCATGATCCCGGTGGACGTTCCC 591
              ||||| ||||| |||   || | |  |||| | | |||   |||||  | | ||   ||
Db      26902 CTGGTCGCCGAACTCGGTGCGGAGGGCGTGCACGCCCGCCCCATCCCCGGCGTCGACACC 26961

Qy        592 GCC---CACTCCCCCCTGATGTACGCCATCGAGGAACGGGTCGTCAGCGGCCTGCTGCCC 648
              |||   ||||| || | | |  || |  | |||| |  |  |  |  ||  ||||
Db      26962 GCCGGGCACTCGCCGCAGGTCGACACGCTGGAGGCCCACCTGCGGAAGGTGCTCGCGCCC 27021

Qy        649 ATCACCCCACGCCCCTCCCGCATCCCCTTCCACTCCTCGGTGACCGGCGGCCGCCTCGAC 708
              || | || ||| ||||| |||||| ||| |||| |||| ||||| || |   |||||
Db      27022 GTCGCGCCCCGCACCTCCGACATCCCGTTCTACTCGACGGTCACCGGAGGACTGATCGAC 27081

Qy        709 ACCCGCGAGCTAGACGCGGCGTACTGGTACCGCAACATGTCGAGCACGGTCCGGTTCGAG 768
              |||  |||||| ||||| |  |||||||||||||||||        ||||   |||||||
Db      27082 ACCGCCGAGCTGGACGCCGACTACTGGTACCGCAACATGCGCGAGCCGGTGGAGTTCGAG 27141

Qy        769 CCCGCCGCCCGGCTGCTTCTGCAGCAGGGGCCCAAGACGTTCGTCGAGATGAGCCCGCAC 828
              |  ||| ||||   ||  |    | || | | |  |||| | |||  ||||||||||
Db      27142 CAGGCCACCCGCGCCCTGATCGCCGACGGCCACGACGTGTTCCTGGAGTCGAGCCCGCAC 27201

Qy        829 CCGGTGCTGACCATGGGCCTCCAGGAGCTCGCCCCGGACCTGGGCGACACCACCGGCACC 888
              ||  ||||| || |  ||||||||||      |  |||  ||   | | | | | |
Db      27202 CCCATGCTGGCCGTCTCCCTCCAGGAGACGATCAGCGACGCCGGTTCCCCGGCGGCCGTC 27261

Qy        889 GCCGACACCGTGATCATGGGCACGCTGCGCCGCGGCCAGGGCACCCTGGACCACTTCC 946
              || |||| ||      ||| |   ||| | ||  |      | | | || || ||
Db      27262 CTCGGCACCCTGCGGCGCGGCCAGGGCGGCCCCCGCTGGCTGGGCGTCGCCCTCTGCC 27319


RESULT 12
AAT80414
ID   AAT80414 standard; DNA; 44377 BP.
XX
```

```
AC   AAT80414;
XX
DT   27-FEB-1998  (first entry)
XX
DE   Platenolide synthase gene cluster.
XX
KW   Tylactone synthase gene cluster; tylG gene; multifunctional protein;
KW   platenolide synthase gene cluster; platenolide production; srmG gene;
KW   polyketide; tylactone synthesis; antibiotic; tylosin; ss.
XX
OS   Streptomyces ambofaciens.
XX
FH   Key             Location/Qualifiers
FT   CDS             350..14002
FT                   /*tag=  a
FT                   /transl_except= (pos:350..352, aa:Met)
FT                   /note= "ORF1 encodes protein shown in AAW22606"
FT   CDS             14046..20036
FT                   /*tag=  b
FT                   /note= "ORF2 encodes protein shown in AAW22607"
FT   CDS             20110..31284
FT                   /*tag=  c
FT                   /transl_except= (pos:20111..20113, aa:Met)
FT                   /note= "ORF3 encodes protein shown in AAW22608"
FT   CDS             31329..36071
FT                   /*tag=  d
FT                   /note= "ORF4 encodes protein shown in AAW22609"
FT   CDS             36155..41830
FT                   /*tag=  e
FT                   /note= "ORF5 encodes protein shown in AAW22610"
XX
PN   EP791655-A2.
XX
PD   27-AUG-1997.
XX
PF   19-FEB-1997;   97EP-0301056.
XX
PR   22-FEB-1996;   96US-0012078.
XX
PA   (ELIL ) LILLY & CO ELI.
XX
PI   Dehoff BS,  Kuhstoss SA,  Rosteck PR,  Sutton KL;
XX
DR   WPI; 1997-418046/39.
DR   P-PSDB; AAW22606-W22610.
XX
PT   DNA encoding Streptomyces fradiae tylactone synthase domain - for
PT   production of tylosin-related polyketide compounds
XX
PS   Example 2; Pages 110-134; 220pp; English.
XX
CC   This sequence represents the platenolide synthase gene cluster of the
CC   invention. This sequence is referred to as the srmG gene, and was
CC   isolated from Streptomyces ambofaciens. This sequence encodes the
CC   multi-functional proteins which direct the synthesis of the polyketide
CC   platenolide. Platenolide is the basic building block of the macrolide
CC   antibiotic spiramycin. This sequence was used along with the tylG gene
CC   (see AAT80413) to create a hybrid ORF1 sequence (see AAT80415). The tylG
CC   gene is the tylactone synthase gene cluster of the invention. The tylG
CC   sequence was isolated from Streptomyces fradiae, and encodes
CC   multifunctional proteins which direct the synthesis of the polyketide
CC   tylactone. Tylactone is the basic building block of the antibiotic
CC   tylosin. The hybrid sequence can be used to transform S. ambofaciens
CC   lacking the srmG ORF1 sequence, or S. fradiae lacking the tylG ORF1
CC   sequence, so that they can produce polyketides. The DNA sequence can be
CC   modified so as to alter the type of carboxylic acids incorporated, the
CC   number of carboxylic acids incorporated and/or the post-condensation
CC   reactions performed, thereby resulting in novel tylosin-related
CC   polyketides.
XX
SQ   Sequence 44377 BP; 4965 A; 15552 C; 17381 G; 6479 T; 0 other;

  Query Match             36.3%;  Score 344;  DB 18;  Length 44377;
  Best Local Similarity   62.5%;  Pred. No. 1.1e-47;
  Matches  599;  Conservative    0;  Mismatches  335;  Indels   24;  Gaps    3;

Qy          1 GTCTTCGTTTTTCCCGGGCAGGGCCCGCAATGGCCGGGCATGGGAAGGGAACTTCT---- 56
              || ||||| || ||||| |||||| |||| ||||| |  ||||      |  || ||
Db      26374 GTGTTCGTCTTCCCCGGCCAGGGCTCGCAGTGGCCCGAGATGGCCGACGGGCTGCTGGCC 26433

Qy         57 -----CGACGCTTCCGACGTCTTCCGGGAGAGCGTCCGCGCCTGCGAAGCCGCGTTCGCG 111
                   |  ||  |||| |  ||||| ||||| || ||||||||||||    |||| ||  |
Db      26434 CGCTCCAGCGGCTCCGGCTCCTTCCTGGAGACCGCCCGCGCCTGCGACCTCGCGCTCCGG 26493

Qy        112 CCCTACGTCGACTGGTCGGTGGAGCAGGTGTTGCGGGACTCGCCGGACGCTCCCGGGCTG 171
              ||| || ||| |||||| ||   | | ||  ||||        || | ||| |||  |||
Db      26494 CCCCACCTCGGCTGGTCCGTCCTGGACGTACTGCGCCGGGAACCCGGCGCGCCCTCGCTC 26553

Qy        172 GACCGGGTGGACGTCGTCCAGCCGACCCTGTTCGCCGTCATGATCTCCCTGGCCGCCCTC 231
              |||||||| ||||| || |||||     |||||| || | ||| |||| || ||||
Db      26554 GACCGGGTCGACGTGGTGCAGCCCGTGCTGTTCACCATGATGGTCTCGCTCGCCGAGACG 26613

Qy        232 TGGCGCTCGCAAGGGGTCGAGCCGTGCGCGGTGCTGGGACACAGCCTGGGCGAGATCGCG 291
              ||||| ||||  || ||||| |||  ||||||  | || |||  || ||||||||||||
Db      26614 TGGCGTTCGCTGGGCGTCGAACCGGCCGCGGTCGTCGGTCACTCCCAGGGCGAGATCGCC 26673

Qy        292 GCAGCCCACGTCTCGGGAGGCCTGTCCCTGGCCGACGCCGCACGCGTGGTGACGCTTTGG 351
              || ||| ||||| | || | |||| | |||| |||||| || ||| | ||  | ||  |
Db      26674 GCCGCCTACGTCGCCGGCGCCCTGACGCTGGACGACGCGGCGCGCATCGTCGCCCTGCGC 26733

Qy        352 AGCCAGGCACAGACCACCCTTGCCGGGACCGGCGCGCTCGTCTCCGTCGCCGCCACGCCG 411
              ||||||||    |      || ||||| |  ||||   | ||| |||| .||    |
Db      26734 AGCCAGGCGTGGCTGCGGCTGGCCGGCAAGGGCGGCATGGTCGCCGTGACCCTGTCCGAA 26793

Qy        412 GATGAGCTCCTGCCCCGAATCGCTCCGTGGACCGAGGACAACCCGGCGCGGCTCGCCGTC 471
                || || |  ||||| | |  || |||| |||                |||||||||||
Db      26794 CGCGACCTGCGTCCCCGCCTGGAGCCCTGGAGCGA------------CCGGCTCGCCGTC 26841

Qy        472 GCAGCCGTCAACGGACCCCGGAGCACAGTCGTTTCCGGTGCCCGCGAGGCCGTCGCGGAC 531
              || |||||||||| |||  || |   | ||| ||||| | ||  || ||| | |||||
Db      26842 GCCGCCGTCAACGGCCCCGAGACCTGCGCCGTCTCCGGGGACCCGGACGCCCTGGCGGAG 26901

Qy        532 CTGGTGGCCGACCTCACCGCCGCGCAGGTGCGCACGCGCATGATCCCGGTGGACGTTCCC 591
              ||||| ||||| ||| |  | |  |||| | | |||  ||||| |  | ||    ||
Db      26902 CTGGTCGCCGAACTCGGTGCGGAGGGCGTGCACGCCCGCCCCATCCCCGGCGTCGACACC 26961

Qy        592 GCC---CACTCCCCCCTGATGTACGCCATCGAGGAACGGGTCGTCAGCGGCCTGCTGCCC 648
              |||   ||||| || | | |  || |  | ||||  |   |    |  |  ||   ||||
Db      26962 GCCGGGCACTCGCCGCAGGTCGACACGCTGGAGGCCCACCTGCGGAAGGTGCTCGCGCCC 27021

Qy        649 ATCACCCCACGCCCCTCCCGCATCCCCTTCCACTCCTCGGTGACCGGCGGCCGCCTCGAC 708
              || | || ||| ||||| |||||| ||| |||| |||| ||||| || |   |||||
Db      27022 GTCGCGCCCCGCACCTCCGACATCCCGTTCTACTCGACGGTCACCGGAGGACTGATCGAC 27081

Qy        709 ACCCGCGAGCTAGACGCGGCGTACTGGTACCGCAACATGTCGAGCACGGTCCGGTTCGAG 768
              |||  |||||| ||||| |  |||||||||||||||||||        ||||  |||||||
Db      27082 ACCGCCGAGCTGGACGCCGACTACTGGTACCGCAACATGCGCGAGCCGGTGGAGTTCGAG 27141

Qy        769 CCCGCCGCCCGGCTGCTTCTGCAGCAGGGGCCCAAGACGTTCGTCGAGATGAGCCCGCAC 828
              |  ||| ||||    ||  |      | || | | |   |||| | |||  ||||||||||
Db      27142 CAGGCCACCCGCGCCCTGATCGCCGACGGCCACGACGTGTTCCTGGAGTCGAGCCCGCAC 27201

Qy        829 CCGGTGCTGACCATGGGCCTCCAGGAGCTCGCCCCGGACCTGGGCGACACCACCGGCACC 888
              ||  ||||| || |   ||||||||||     |   |||   ||   | |  | | |  |
Db      27202 CCCATGCTGGCCGTCTCCCTCCAGGAGACGATCAGCGACGCCGGTTCCCCGGCGGCCGTC 27261

Qy        889 GCCGACACCGTGATCATGGGCACGCTGCGCCGCGGCCAGGGCACCCTGGACCACTTCC 946
              || |||| ||       ||| |  ||| | || |     | | | || || ||
Db      27262 CTCGGCACCCTGCGGCGCGGCCAGGGCGGCCCCCGCTGGCTGGGCGTCGCCCTCTGCC 27319


RESULT 13
AAV99255
ID   AAV99255 standard; DNA; 1010 BP.
XX
AC   AAV99255;
```

XX
DT 09-MAR-1999 (first entry)
XX
DE DNA encoding an active acyltransferase domain (NidAT5) from a PKS.
XX
KW Enzymatically active acyltransferase domain; polyketide synthase; PKS;
KW polyketide synthesis; antimicrobial; anticancer; antifungal;
KW immunosuppressant; anthelmintic; NidAT5; ethyl acyltransferase domain;
KW module 5; PKS cluster; ds.
XX
OS Streptomyces caelestis.
XX
PN WO9851695-A2.
XX
PD 19-NOV-1998.
XX
PF 13-MAY-1998; 98WO-US09518.
XX
PR 16-MAY-1997; 97US-0858003.
XX
PA (ABBO) ABBOTT LAB.
XX
PI Kakavas SJ, Katz L, Pereda-Lopez A, Ruan X, Stassi DL;
PI Summers RG;
XX
DR WPI; 1999-070114/06.
DR P-PSDB; AAW87714.
XX
PT New polyketides produced by microorganism having altered polyketide
PT synthase genes - especially chimeras containing a heterologous
PT acyltransferase domain, useful as e.g. antimicrobials, anticancer
PT agents
XX
PS Claim 24; Fig 21; 137pp; English.
XX
CC The present sequence encodes an enzymatically active acyltransferase
CC domain, NidAT5, from a polyketide synthase (PKS). The NidAT5 protein
CC is derived from the ethyl acyltransferase domain from module 5 of the
CC PKS cluster of Streptomyces caelestis NRRL-2821. The sequence
CC can be used to replace acyltransferase domains of other PKS enzymes,
CC which are then used to generate novel polyketides by de novo biosynthesis
CC rather than by chemical modification. Polyketides of the invention have
CC the methyl groups on the macrolactone ring substituted with -H, -Et
CC and/or -OH, and the ethyl side chain substituted with hydroxymethyl or
CC dihydroxycyclohexylmethyl side chains. Polyketides of the invention are
CC antimicrobial, anticancer, antifungal, immunosuppressant and/or
CC anthelmintic agents, or their intermediates.
XX
SQ Sequence 1010 BP; 104 A; 353 C; 402 G; 151 T; 0 other;

Query Match 35.6%; Score 337.4; DB 20; Length 1010;
Best Local Similarity 61.4%; Pred. No. 2e-46;
Matches 583; Conservative 0; Mismatches 351; Indels 15; Gaps 2;

```
Qy          1 GTCTTCGTTTTTCCCGGGCAGGGCCCGCAATGGCCGGGCATGGGAAGGGAACTTCTCGAC 60
              || ||||| || ||||| |||||| |||| ||| | || ||||    ||  || || ||
Db         13 GTGTTCGTGTTCCCCGGCCAGGGCTCGCAGTGGGCCGGAATGGCCGAGGGGCTGCTGGAG 72

Qy         61 GCTTCCGACGTCTTCCGGGAGAGCGTCCGCGCCTGCGAAGCCGCGTTCGCGCCCTACGTC 120
                ||||  ||  ||||||      | |  | | ||||| |||||| |   ||| ||| ||
Db         73 CGGTCCGGCGCGTTCCGGAGTGCGGCCGACTCGTGCGACGCCGCGCTGCGGCCGTACCTC 132

Qy        121 GACTGGTCGGTGGAGCAGGTGTTGCGGGACTCGCCGGACGCTCCCGGGCTGGACCGGGTG 180
              | |||||||||  |   ||| |||| |     |||||||| |||  ||| ||||||||
Db        133 GGCTGGTCGGTGCTGAGCGTGCTGCGCGGGGAACCGGACGCGCCCTCGCTCGACCGGGTC 192

Qy        181 GACGTCGTCCAGCCGACCCTGTTCGCCGTCATGATCTCCCTGGCCGCCCTCTGGCGCTCG 240
              ||||||||  ||||||   |||||| |  | ||| |||| || || ||  |||||||  ||
Db        193 GACGTCGTGCAGCCGGTGCTGTTCACGATGATGGTCTCGCTCGCGGCGGTCTGGCGTGCG 252

Qy        241 CAAGGGGTCGAGCCGTGCGCGGTGCTGGGACACAGCCTGGGCGAGATCGCGGCAGCCCAC 300
              |  ||||| || |||   |||||  | || |||    | ||| |||||||| || |||||
Db        253 CTGGGGGTGGAACCGGCGGCGGTCGTCGGGCACTCGCAGGGTGAGATCGCCGCTGCCCAT 312

Qy        301 GTCTCGGGAGGCCTGTCCCTGGCCGACGCCGCACGCGTGGTGACGCTTTGGAGCCAGGCA 360
              ||| | || |  ||||| |||| |||| | || ||  | ||  | ||  | || | |||
Db        313 GTCGCCGGTGCGCTGTCGCTGGACGACTCGGCCCGGATCGTCGCCCTGCGCAGTCGGGCG 372

Qy        361 CAGACCACCCTTGCCGGGACCGGCGCGCTCGTCTCCGTCGCCGCCACGCCGGATGAGCTC 420
                |  |   || || || |  |||   | ||  | ||  |      || |||| |||||
Db        373 TGGCTCGGACTGGCGGGCAAGGGCGGCATGGTGGCGGTGCCGATGCCGGCGGAGGAGCTG 432

Qy        421 CTGCCCCGAATCGCTCCGTGGACCGAGGACAACCCGGCGCGGCTCGCCGTCGCAGCCGTC 480
              | ||| ||  | |  |||||                ||  || || |||||||| ||||||
Db        433 CGGCCGCGGCTGGTGACGTGG-----------GGGGACCGTCTGGCCGTCGCCGCCGTC 480

Qy        481 AACGGACCCCGGAGCACAGTCGTTTCCGGTGCCCGCGAGGCCGTCGCGGACCTGGTGGCC 540
              ||| | ||| |   |   | |||  | || | ||  |||||  | || || ||||||||
Db        481 AACAGCCCCGGTTCCTGCGCCGTCGCAGGCGACCCGGAGGCGCTGGCCGAACTGGTGGCG 540

Qy        541 GACCTCACCGCCGCGCAGGTGCGCACGCGCATGATCCCGGTGGACGTTCCCGC---CCAC 597
                || ||||  | |  |||| | | ||   |||||| |  | ||   | ||   ||||
Db        541 CTGCTGACCGGTGAGGGGGTGCACGCCCGGCCGATCCCCGGCGTCGACACGGCGGGCCAC 600

Qy        598 TCCCCCCTGATGTACGCCATCGAGGAACGGGTCGTCAGCGGCCTGCTGCCCATCACCCCA 657
              || || | | || |||   |   ||  |   |   |     | |||   ||  || ||||
Db        601 TCGCCGCAGGTGGACGCGTTGCGGGCTCATCTGCTGGAGGTGCTGGCCCCGGTCGCCCCC 660

Qy        658 CGCCCCTCCCGCATCCCCTTCCACTCCTCGGTGACCGGCGGCCGCCTCGACACCCGCGAG 717
              || ||  ||  |||||| ||| ||||  |||||||||||||| |  || |||  |  ||||
Db        661 CGACCGGCCGACATCCCGTTCTACTCGACGGTGACCGGCGGGCTGCTGGACGGCACCGAG 720

Qy        718 CTAGACGCGGCGTACTGGTACCGCAACATGTCGAGCACGGTCCGGTTCGAGCCCGCCGCC 777
              || |||||| |||||||||||||||||||        | |||  |||||||  ||| |
Db        721 CTGGACGCGACGTACTGGTACCGCAACATGCGCGAGCCCGTCGAGTTCGAGCGGGCCACA 780

Qy        778 CGGCTGCTTCTGCAGCAGGGGCCCAAGACGTTCGTCGAGATGAGCCCGCACCCGGTGCTG 837
              |||  |||  |     | |||| | |    ||| | |||| ||||||||| ||  |||||
Db        781 CGGGCGCTGATCGCCGACGGGCACGACGTCTTCCTGGAGACGAGCCCGCATCCCATGCTG 840

Qy        838 ACCATGGGCCTCCAGGAGCTCGCCCCGGACCTGGGCGACACCACCGGCACCGCCGACACC 897
              || |||  ||  || ||  | | | |||   |||  |  | | |        ||  |||
Db        841 GCCGTGGCGCTGGAGCAGACGGTCACCGACGCCGGCACCGACGCGGCGGTGCTCGGGACC 900

Qy        898 GTGATCATGGGCACGCTGCGCCGCGGCCAGGGCACCCTGGACCACTTCC 946
              ||  |  ||     | |  ||  |   |||||  |  || ||
Db        901 CTGCGCCGCCGCCACGGCGGTCCTCGCGCGCTGGCCCTGGCCGTCTGCC 949
```

RESULT 14
AAA72521
ID AAA72521 standard; DNA; 1010 BP.
XX
AC AAA72521;
XX
DT 20-NOV-2000 (first entry)
XX
DE DNA sequence of the ethyl AT module of the niddamycin PKS cluster.
XX
KW Erythromycin derivative; polyketide synthase; PKS; acetyltransferase;
KW antibiotic; antifungal; anticancer; immunosuppressant; antihelminithic;
KW ethyl AT; niddamycin; ds.
XX
OS Streptomyces caelestis.
XX
PN US6063561-A.
XX
PD 16-MAY-2000.
XX
PF 13-MAY-1998; 98US-0078166.
XX
PR 17-JAN-1991; 91US-0642734.
PR 16-MAY-1997; 97US-0858003.
XX
PA (ABBO) ABBOTT LAB.
XX

```
PI   Katz L,  Stassi DL,  Ruan X,  Pereda-Lopez A,  Kakavas SJ,  Summers RG;
XX
DR   WPI; 2000-386760/33.
DR   P-PSDB; AAB12926.
XX
PT   New erythromycin derivatives produced by microorganisms with altered
PT   polyketide synthase genes  -
XX
PS   Example 20; Fig 21; 94pp; English.
XX
CC   This invention relates to novel erythromycin derivatives, in which the
CC   methyl groups on the macrolactone ring have been substituted  with H, Et
CC   and/or OH and the ethyl side chain has been substituted with
CC   hydroxymethyl or dihydroxycyclohexylmethyl side chains. Included in the
CC   invention are vectors containing polynucleotide sequences which encode an
CC   enzymatically active acetyltransferase domain from Streptomyces. Also
CC   included are methods for altering the substrate specificity of a
CC   polyketide synthase in a microorganism using the vectors of the
CC   invention. Polyketides are are a large class of natural products
CC   including antibiotics, antifungal, anticancer and immunosuppressant
CC   compounds. Polyketide synthase (PKS) is a complex of enzyme activities
CC   responsible for the biosynthesis of polyketides. The erythromycin
CC   derivatives exhibit antibiotic, antifungal, anticancer, antihelminithic
CC   and immunosuppressant activity.
CC   The present sequence represents the DNA sequence encoding the ethyl
CC   acetyltransferase (AT) module of the niddamycin PKS cluster isolated from
CC   Streptomyces caelestis. The sequence is used in the production of a
CC   vector of the invention.
XX
SQ   Sequence 1010 BP; 104 A; 353 C; 402 G; 151 T; 0 other;

  Query Match             35.6%;  Score 337.4;  DB 21;  Length 1010;
  Best Local Similarity   61.4%;  Pred. No. 2e-46;
  Matches  583;  Conservative    0;  Mismatches  351;  Indels   15;  Gaps    2;

Qy          1 GTCTTCGTTTTTCCCGGGCAGGGCCCGCAATGGCCGGGCATGGGAAGGGAACTTCTCGAC 60
              || ||||| || ||||| |||||| |||| ||| | || ||||    ||  || || ||
Db         13 GTGTTCGTGTTCCCCGGCCAGGGCTCGCAGTGGGCCGGAATGGCCGAGGGGCTGCTGGAG 72

Qy         61 GCTTCCGACGTCTTCCGGGAGAGCGTCCGCGCCTGCGAAGCCGCGTTCGCGCCCTACGTC 120
                |||| ||  ||||||      | |  | | ||||| ||||||  |   ||| ||| ||
Db         73 CGGTCCGGCGCGTTCCGGAGTGCGGCCGACTCGTGCGACGCCGCGCTGCGGCCGTACCTC 132

Qy        121 GACTGGTCGGTGGAGCAGGTGTTGCGGGACTCGCCGGACGCTCCCGGGCTGGACCGGGTG 180
              | ||||||||||  |   ||| |||| |     |||||||| |||  ||| ||||||||
Db        133 GGCTGGTCGGTGCTGAGCGTGCTGCGCGGGGAACCGGACGCGCCCTCGCTCGACCGGGTC 192

Qy        181 GACGTCGTCCAGCCGACCCTGTTCGCCGTCATGATCTCCCTGGCCGCCCTCTGGCGCTCG 240
              |||||||| ||||||   ||||||  |  | ||| |||| || || ||  |||||||  ||
Db        193 GACGTCGTGCAGCCGGTGCTGTTCACGATGATGGTCTCGCTCGCGGCGGTCTGGCGTGCG 252

Qy        241 CAAGGGGTCGAGCCGTGCGCGGTGCTGGGACACAGCCTGGGCGAGATCGCGGCAGCCCAC 300
              |  ||||| || |||   |||||  | || |||    | ||| ||||||||  || |||||
Db        253 CTGGGGGTGGAACCGGCGGCGGTCGTCGGGCACTCGCAGGGTGAGATCGCCGCTGCCCAT 312

Qy        301 GTCTCGGGAGGCCTGTCCCTGGCCGACGCCGCACGCGTGGTGACGCTTTGGAGCCAGGCA 360
              ||| | || |  ||||| |||| |||| | || ||  | ||  | ||  | || | |||
Db        313 GTCGCCGGTGCGCTGTCGCTGGACGACTCGGCCCGGATCGTCGCCCTGCGCAGTCGGGCG 372

Qy        361 CAGACCACCCTTGCCGGGACCGGCGCGCTCGTCTCCGTCGCCGCCACGCCGGATGAGCTC 420
                |   || || || |  ||||   | ||  | ||  |     || |||| |||||
Db        373 TGGCTCGGACTGGCGGGCAAGGGCGGCATGGTGGCGGTGCCGATGCCGGCGGAGGAGCTG 432

Qy        421 CTGCCCCGAATCGCTCCGTGGACCGAGGACAACCCGGCGCGGCTCGCCGTCGCAGCCGTC 480
              | ||| ||  | |   |||||              ||  || || |||||||| ||||||
Db        433 CGGCCGCGGCTGGTGACGTGG------------GGGGACCGTCTGGCCGTCGCCGCCGTC 480

Qy        481 AACGGACCCCGGAGCACAGTCGTTTCCGGTGCCCGCGAGGCCGTCGCGGACCTGGTGGCC 540
              ||| | ||| |   |  | |||  | || | ||  |||||  | || || ||||||||
Db        481 AACAGCCCCGGTTCCTGCGCCGTCGCAGGCGACCCGGAGGCGCTGGCCGAACTGGTGGCG 540

Qy        541 GACCTCACCGCCGCGCAGGTGCGCACGCGCATGATCCCGGTGGACGTTCCCGC---CCAC 597
                || ||||  | |  ||||| | | ||   |||||| |  | ||   | ||   ||||
Db        541 CTGCTGACCGGTGAGGGGGTGCACGCCCGGCCGATCCCCGGCGTCGACACGGCGGGCCAC 600

Qy        598 TCCCCCCTGATGTACGCCATCGAGGAACGGGTCGTCAGCGGCCTGCTGCCCATCACCCCA 657
              || || | | || ||||  |   ||  |   |  |     | |||   ||  || ||||
Db        601 TCGCCGCAGGTGGACGCGTTGCGGGCTCATCTGCTGGAGGTGCTGGCCCCGGTCGCCCCC 660

Qy        658 CGCCCCTCCCGCATCCCCTTCCACTCCTCGGTGACCGGCGGCCGCCTCGACACCCGCGAG 717
              || ||  ||  |||||| ||| ||||  |||||||||||||  |  || |||  |  ||||
Db        661 CGACCGGCCGACATCCCGTTCTACTCGACGGTGACCGGCGGGCTGCTGGACGGCACCGAG 720

Qy        718 CTAGACGCGGCGTACTGGTACCGCAACATGTCGAGCACGGTCCGGTTCGAGCCCGCCGCC 777
              || |||||| |||||||||||||||||||        | |||  ||||||||  ||| |
Db        721 CTGGACGCGACGTACTGGTACCGCAACATGCGCGAGCCCGTCGAGTTCGAGCGGGCCACA 780

Qy        778 CGGCTGCTTCTGCAGCAGGGGCCCAAGACGTTCGTCGAGATGAGCCCGCACCCGGTGCTG 837
              |||  |||  |     | |||| | |    ||| | |||| |||||||| ||  |||||
Db        781 CGGGCGCTGATCGCCGACGGGCACGACGTCTTCCTGGAGACGAGCCCGCATCCCATGCTG 840

Qy        838 ACCATGGGCCTCCAGGAGCTCGCCCCGGACCTGGGCGACACCACCGGCACCGCCGACACC 897
              || |||  ||  || ||  | | | |||   |||  |  | | |         ||  |||
Db        841 GCCGTGGCGCTGGAGCAGACGGTCACCGACGCCGGCACCGACGCGGCGGTGCTCGGGACC 900

Qy        898 GTGATCATGGGCACGCTGCGCCGCGGCCAGGGCACCCTGGACCACTTCC 946
                ||  |   ||     | |   ||  |    |||||| |  || ||
Db        901 CTGCGCCGCCGCCACGGCGGTCCTCGCGCGCTGGCCCTGGCCGTCTGCC 949


RESULT 15
AAT80413
ID   AAT80413 standard; DNA; 43280 BP.
XX
AC   AAT80413;
XX
DT   27-FEB-1998  (first entry)
XX
DE   Tylactone synthase gene cluster.
XX
KW   Tylactone synthase gene cluster; tylG gene; multifunctional protein;
KW   polyketide; tylactone synthesis; antibiotic; tylosin; ss.
XX
OS   Streptomyces fradiae.
XX
FH   Key             Location/Qualifiers
FT   CDS             816..14243
FT                   /*tag=  a
FT                   /transl_except= (pos: 816..818, aa: Met)
FT                   /note= "ORF1 encodes protein shown in AAW22601"
FT   CDS             14351..19945
FT                   /*tag=  b
FT                   /transl_except= (pos: 14351..14353, aa: Met)
FT                   /note= "ORF2 encodes protein shown in AAW22602"
FT   CDS             20010..31199
FT                   /*tag=  c
FT                   /transl_except= (pos: 20010..20012, aa: Met)
FT                   /note= "ORF3 encodes protein shown in AAW22603"
FT   CDS             31232..36067
FT                   /*tag=  d
FT                   /note= "ORF4 encodes protein shown in AAW22604"
FT   CDS             36249..41774
FT                   /*tag=  e
FT                   /note= "ORF5 encodes protein shown in AAW22605"
XX
PN   EP791655-A2.
XX
PD   27-AUG-1997.
XX
PF   19-FEB-1997;   97EP-0301056.
XX
PR   22-FEB-1996;   96US-0012078.
XX
PA   (ELIL ) LILLY & CO ELI.
XX
PI   Dehoff BS,  Kuhstoss SA,  Rosteck PR,  Sutton KL;
```

XX
DR WPI; 1997-418046/39.
DR P-PSDB; AAW22601-W22605.
XX
PT DNA encoding Streptomyces fradiae tylactone synthase domain - for
PT production of tylosin-related polyketide compounds
XX
PS Claim 2; Pages 8-66; 220pp; English.
XX
CC This sequence represents the tylactone synthase gene cluster of the
CC invention. This sequence is also referred to as the tylG gene, and was
CC isolated from Streptomyces fradiae. This sequence encodes multifunctional
CC proteins which direct the synthesis of the polyketide tylactone, isolated
CC from Streptomyces fradiae. Tylactone is the basic building block of the
CC antibiotic tylosin. The DNA sequence can be modified so as to alter the
CC type of carboxylic acids incorporated, the number of carboxylic acids
CC incorporated and/or the post-condensation reactions performed, thereby
CC resulting in novel tylosin-related polyketides.
XX
SQ Sequence 43280 BP; 5269 A; 15187 C; 16651 G; 6173 T; 0 other;

Query Match 33.7%; Score 319.2; DB 18; Length 43280;
Best Local Similarity 60.2%; Pred. No. 1.2e-43;
Matches 572; Conservative 0; Mismatches 363; Indels 15; Gaps 2;

```
Qy          1 GTCTTCGTTTTTCCCGGGCAGGGCCCGCAATGGCCGGGCATGGGAAGGGAACTTCTCGAC 60
              |||||||| || |||||||||||| |||| ||||||    ||||    |||| || ||||||
Db      26295 GTCTTCGTCTTCCCCGGGCAGGGCTCGCAGTGGCCGTCGATGGCCCGGGACCTGCTCGAC 26354

Qy         61 GCTTCCGACGTCTTCCGGGAGAGCGTCCGCGCCTGCGAAGCCGCGTTCGCGCCCTACGTC 120
                |   || |||||| ||||  |      |||||||| |||||| |       | |  |
Db      26355 CGCGCGCCCGCCTTCCGCGAGACGGCGAAGGCCTGCGACGCCGCGCTGAGCGTCCATCTG 26414

Qy        121 GACTGGTCGGTGGAGCAGGTGTTGCGGGACTCGCCGGACGCTCCCGGGCTGGACCGGGTG 180
              |||||||| |||    | ||  | | |||   |||||||||  ||   ||||  ||||||
Db      26415 GACTGGTCCGTGCTCGATGTCCTCCAGGAGAAGCCGGACGCGCCGCCGCTGAGCCGGGTC 26474

Qy        181 GACGTCGTCCAGCCGACCCTGTTCGCCGTCATGATCTCCCTGGCCGCCCTCTGGCGCTCG 240
              ||||| || |||||     ||||||  |   | ||| | || || ||||||   ||||||
Db      26475 GACGTGGTGCAGCCCGTGCTGTTCACGATGATGCTGTCGCTCGCCGCCTGCTGGCGGGAC 26534

Qy        241 CAAGGGGTCGAGCCGTGCGCGGTGCTGGGACACAGCCTGGGCGAGATCGCGGCAGCCCAC 300
              |  || ||| | |||  ||| ||| |||| |||  || ||| ||||||||||| |||  |
Db      26535 CTCGGCGTCCACCCGGCCGCCGTGGTGGGCCACTCCCAGGGAGAGATCGCGGCGGCCTGC 26594

Qy        301 GTCTCGGGAGGCCTGTCCCTGGCCGACGCCGCACGCGTGGTGACGCTTTGGAGCCAGGCA 360
              ||  | || |  || |||||||  ||||| || ||  | ||| ||||  | |||| ||||
Db      26595 GTGGCCGGCGCGCTCTCCCTGGAGGACGCGGCGCGGATCGTGGCGCTGCGCAGCCGGGCA 26654

Qy        361 CAGACCACCCTTGCCGGGACCGGCGCGCTCGTCTCCGTCGCCGCCACGCCGGATGAGCTC 420
                | ||| || ||||| |  ||||    | | | ||||| ||     ||   |
Db      26655 TGGCTCACACTGGCCGGCAAGGGCGGCATGGCCGCCGTCTCCCTGCCGGAAG-------- 26706

Qy        421 CTGCCCCGAATCGCTCCGTGGACCGAGGACAACCCGGCGCGGCTCGCCGTCGCAGCCGTC 480
                |||| |      | ||| ||||      |  |  ||||||  | || || || ||
Db      26707 ----CCCGGCTGCGCGAGCGGATCGAGCGGTTCGGGCAGCGGCTGTCGGTGGCCGCGGTG 26762

Qy        481 AACGGACCCCGGAGCACAGTCGTTTCCGGTGCCCGCGAGGCCGTCGCGGACCTGGTGGCC 540
              ||| | ||  | |   | |  ||  |||||| |    || ||   |   ||| ||| ||||
Db      26763 AACAGCCCGGGCACGGCGGCGGTCGCCGGTGACGTGGACGCGCTGCGGGAACTGCTGGCG 26822

Qy        541 GACCTCACCGCCGCGCAGGTGCG---CACGCGCATGATCCCGGTGGACGTTCCCGCCCAC 597
              || || ||||| | |     | ||   || ||  ||    |   ||||||     ||| ||||
Db      26823 GAGCTGACCGCGGAGGGCATCCGGGCCAAGCCGATCCCCGGCGTGGACACGGCCGGCCAC 26882

Qy        598 TCCCCCCTGATGTACGCCATCGAGGAACGGGTCGTCAGCGGCCTGCTGCCCATCACCCCA 657
              ||| | | | || ||| | |  |||| |   || ||   |  |||  |||  || ||||
Db      26883 TCCGCGCAGGTGGACGGCCTGAAGGAGCATCTCTTCGAGGTGCTGGCGCCGGTCTCCCCG 26942

Qy        658 CGCCCCTCCCGCATCCCCTTCCACTCCTCGGTGACCGGCGGCCGCCTCGACACCCGCGAG 717
              ||| ||||  |||||| ||| ||||  |||||||| ||||  |  || ||||||     |
Db      26943 CGCTCCTCGGACATCCCGTTCTACTCGACGGTGACGGGCGCGCCGCTGGACACCGAGCGG 27002

Qy        718 CTAGACGCGGCGTACTGGTACCGCAACATGTCGAGCACGGTCCGGTTCGAGCCCGCCGCC 777
              || ||||| | |||||||||||||||||||  |    | ||   |||||||   |||| |
Db      27003 CTGGACGCCGGGTACTGGTACCGCAACATGCGGGAGCCCGTGGAGTTCGAGAAGGCCGTC 27062

Qy        778 CGGCTGCTTCTGCAGCAGGGGCCCAAGACGTTCGTCGAGATGAGCCCGCACCCGGTGCTG 837
              ||   ||  |      | ||   | |   |||| | |||   | ||||||||| ||||
Db      27063 AGGGCACTGATCGCCGACGGCTACGACCTGTTCCTGGAGTGCAACCCGCACCCGATGCTC 27122

Qy        838 ACCATGGGCCTCCAGGAGCTCGCCCCGGACCTGGGCGACACCACCGGCACCGCCGACACC 897
              |||||   ||  | |||    | | |||   |||| |  |  |  |        |||||
Db      27123 GCCATGTCGCTGGACGAGACACTCACCGACAGCGGCGGCCACGGCACCGTGATGCACACC 27182

Qy        898 GTGATCATGGGCACGCTGCGCCGCGGCCAGGGCACCCTGGACCACTTCCT 947
              |   |  |   | |   ||  |   |   |  | |||  | || |||
Db      27183 CTCCGCCGGCAGAAGGGCAGCGCCAAGGACTTCGGCATGGCGCTCTGCCT 27232
```

Search completed: June 17, 2003, 13:47:58
Job time : 207.674 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 17, 2003, 11:45:13 ; Search time 41.0664 Seconds
(without alignments)
7079.508 Million cell updates/sec

Title: US-09-914-286-1_COPY_85_1032
Perfect score: 948
Sequence: 1 gtcttcgtttttcccgggca.....gcaccctggaccacttcctg 948

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	348.2	36.7	28958	1	US-08-258-261B-6	Sequence 6, Appli
2	348.2	36.7	28958	1	US-08-456-837-6	Sequence 6, Appli
3	348.2	36.7	28958	1	US-08-457-342-6	Sequence 6, Appli
4	348.2	36.7	28958	1	US-08-457-646A-6	Sequence 6, Appli
5	348.2	36.7	28958	1	US-08-458-076A-6	Sequence 6, Appli
6	348.2	36.7	28958	1	US-08-764-233A-4	Sequence 4, Appli
7	348.2	36.7	28958	1	US-08-457-335A-6	Sequence 6, Appli
8	348.2	36.7	28958	1	US-08-729-214-6	Sequence 6, Appli
9	348.2	36.7	28958	3	US-09-028-934-6	Sequence 6, Appli
10	348.2	36.7	49377	1	US-08-764-233A-1	Sequence 1, Appli
11	344	36.3	44377	2	US-08-804-227C-7	Sequence 7, Appli
12	344	36.3	44377	2	US-08-804-198-1	Sequence 1, Appli
13	337.4	35.6	1010	3	US-08-858-003-29	Sequence 29, Appl
14	337.4	35.6	1010	3	US-09-078-166-29	Sequence 29, Appl
15	337.4	35.6	1010	4	US-08-997-467-29	Sequence 29, Appl
16	319.2	33.7	43280	2	US-08-804-227C-1	Sequence 1, Appli
17	318.4	33.6	20235	1	US-07-642-734C-3	Sequence 3, Appli
18	318.4	33.6	20235	3	US-08-439-009A-3	Sequence 3, Appli
19	300.6	31.7	80161	3	US-09-036-987A-1	Sequence 1, Appli
20	300.6	31.7	80161	4	US-09-370-700-1	Sequence 1, Appli
c 21	293.6	31.0	4403765	4	US-09-103-840A-2	Sequence 2, Appli
c 22	293.6	31.0	4411529	4	US-09-103-840A-1	Sequence 1, Appli
23	284	30.0	13842	4	US-09-105-537-30	Sequence 30, Appl
24	284	30.0	36778	4	US-09-105-537-5	Sequence 5, Appli
25	284	30.0	38506	3	US-09-320-878-19	Sequence 19, Appl
c 26	283.2	29.9	1881	4	US-09-434-288-5	Sequence 5, Appli
27	279.6	29.5	1035	3	US-08-858-003-30	Sequence 30, Appl
28	279.6	29.5	1035	3	US-09-078-166-30	Sequence 30, Appl
29	279.6	29.5	1035	4	US-08-997-467-30	Sequence 30, Appl
30	267.4	28.2	11219	1	US-07-642-734C-1	Sequence 1, Appli
31	267.4	28.2	11219	3	US-08-439-009A-1	Sequence 1, Appli
32	267.2	28.2	11220	4	US-09-105-537-32	Sequence 32, Appl
33	262.6	27.7	4689	4	US-09-105-537-34	Sequence 34, Appl
34	253.2	26.7	50937	4	US-09-428-517-1	Sequence 1, Appli
35	250.6	26.4	4041	4	US-09-105-537-36	Sequence 36, Appl
36	243	25.6	4411529	4	US-09-103-840A-1	Sequence 1, Appli
37	233.2	24.6	33529	4	US-09-144-085-3	Sequence 3, Appli
38	232.6	24.5	4403765	4	US-09-103-840A-2	Sequence 2, Appli
39	231	24.4	68750	3	US-09-335-409-1	Sequence 1, Appli
40	231	24.4	68750	4	US-09-568-102-1	Sequence 1, Appli
41	231	24.4	68750	4	US-09-567-969-1	Sequence 1, Appli
42	231	24.4	68750	4	US-09-568-480-1	Sequence 1, Appli
43	231	24.4	68750	4	US-09-568-486-1	Sequence 1, Appli
44	231	24.4	68750	4	US-09-568-472-1	Sequence 1, Appli
45	231	24.4	68750	4	US-09-567-899-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1
US-08-258-261B-6
; Sequence 6, Application US/08258261B
; Patent No. 5639949
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James M.
; APPLICANT: Beck, James Joseph
; APPLICANT: Hill, Dwight Steven
; APPLICANT: Ryals, John Andrew
; APPLICANT: Gaffney, Thomas Deane
; APPLICANT: Lam, Stephen Ting
; APPLICANT: Hammer, Phillip E.
; APPLICANT: Uknes, Scott Joseph
; TITLE OF INVENTION: Genes for the synthesis of
; TITLE OF INVENTION: antipathogenic substances
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ciba-Geigy Corporation
; STREET: 7 Skyline Drive
; CITY: Hawthorne
; STATE: NY
; COUNTRY: USA
; ZIP: 10532
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/258,261B
; FILING DATE: 08-JUN-1994
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/457,205
; FILING DATE: 01-JUN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Elmer, James Scott
; REGISTRATION NUMBER: 36,129
; REFERENCE/DOCKET NUMBER: CGC 1506/CIP3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8614
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 28958 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)

```
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
US-08-258-261B-6

  Query Match            36.7%;  Score 348.2;  DB 1;  Length 28958;
  Best Local Similarity  62.7%;  Pred. No. 4.5e-57;
  Matches  564;  Conservative    0;  Mismatches  323;  Indels   12;  Gaps    1;

Qy          1 GTCTTCGTTTTTCCCGGGCAGGGCCCGCAATGGCCGGGCATGGGAAGGGAACTTCTCGAC 60
              |||||||| ||||| ||||| ||| |||| |||   || ||||         || ||||||
Db       9940 GTCTTCGTCTTTCCTGGGCAAGGCTCGCAGTGGGAAGGGATGGCCCTCTCCCTGCTCGAC 9999

Qy         61 GCTTCCGACGTCTTCCGGGAGAGCGTCCGCGCCTGCGAAGCCGCGTTCGCGCCCTACGTC 120
              | ||    |||||||| |       ||    || |||||   |||| |||| ||  |||||
Db      10000 TCCTCGCCGGTCTTCCGCGCTCAGCTCGAAGCATGCGAGCGCGCGCTCGCTCCTCACGTC 10059

Qy        121 GACTGGTCGGTGGAGCAGGTGTTGCGGGACTCGCCGGACGCTCCCGGGCTGGACCGGGTG 180
              || |||    ||       ||  ||||   |     || ||| |||   || ||||| ||
Db      10060 GAGTGGAGCCTGCTCGCCGTCCTGCGCCGCGACGAGGGCGCCCCCTCCCTCGACCGCGTC 10119

Qy        181 GACGTCGTCCAGCCGACCCTGTTCGCCGTCATGATCTCCCTGGCCGCCCTCTGGCGCTCG 240
              |||||||| |||||  |||| || ||||||||| ||||||||||||||||||||||||||
Db      10120 GACGTCGTACAGCCCGCCCTCTTTGCCGTCATGGTCTCCCTGGCCGCCCTCTGGCGCTCG 10179

Qy        241 CAAGGGGTCGAGCCGTGCGCGGTGCTGGGACACAGCCTGGGCGAGATCGCGGCAGCCCAC 300
              |  || |||||||||    ||| ||  | || |||||||  |||||||||||| || |||  |
Db      10180 CTCGGCGTCGAGCCCGCCGCCGTCGTCGGCCACAGCCAGGGCGAGATCGCCGCCGCCTTC 10239

Qy        301 GTCTCGGGAGGCCTGTCCCTGGCCGACGCCGCACGCGTGGTGACGCTTTGGAGCCAGGCA 360
              ||| | || |  || ||||| |  ||||| || ||| | |    | ||  | ||  | |||
Db      10240 GTCGCAGGCGCTCTCTCCCTCGAGGACGCGGCGCGCATCGCCGCCCTGCGCAGGAAAGCG 10299

Qy        361 CAGACCACCCTTGCCGGGACCGGCGCGCTCGTCTCCGTCGCCGCCACGCCGGATGAGCTC 420
              |  |||||| | | ||| | |||||    | | | ||||||     |    |    || |||
Db      10300 CTCACCACCGTCGGCGGCAACGGCGGCATGGCCGCCGTCGAGCTCGGCGCCTCCGACCTC 10359

Qy        421 CTGCCCCGAATCGCTCCGTGGACCGAGGACAACCCGGCGCGGCTCGCCGTCGCAGCCGTC 480
              | | ||    ||||||| |||  |||                ||||| ||  ||| ||||||
Db      10360 CAGACCTACCTCGCTCCCTGGGGCGA----------CAGGCTCTCCACCGCCGCCGTC 10407

Qy        481 AACGGACCCCGGAGCACAGTCGTTTCCGGTGCCCGCGAGGCCGTCGCGGACCTGGTGGCC 540
              ||| | ||| ||   ||  |||| ||||| |  | ||  |||||||  |  ||| | | |
Db      10408 AACAGCCCCAGGGCTACCCTCGTATCCGGCGAGCCCGCCGCCGTCGACGCGCTGCTCGAC 10467

Qy        541 GACCTCACCGCCGCGCAGGTGCGCACGCGCATGATCCCGGTGGACGTTCCCGCCCACTCC 600
              | |||||||||| |  ||||  | | |||| ||||  || |||    || ||||||||
Db      10468 GTCCTCACCGCCACCAAGGTGTTCGCCCGCAAGATCCGCGTCGACTACGCCTCCCACTCC 10527

Qy        601 CCCCTGATGTACGCCATCGAGGAACGGGTCGTCAGCGGCCTGCTGCCCATCACCCCACGC 660
              |||| |||| ||||| || | ||   | ||| |   || ||       |||| | || ||
Db      10528 GCCCAGATGGACGCCGTCCAAGACGAGCTCGCCGCAGGTCTAGCCAACATCGCTCCTCGG 10587

Qy        661 CCCTCCCGCATCCCCTTCCACTCCTCGGTGACCGGCGGCCGCCTCGACACCCGCGAGCTA 720
              | | |    ||||  |  | ||  | ||| ||||||  | | ||||||  |  ||||||
Db      10588 ACGTGCGAGCTCCCTCTTTATTCGACCGTCACCGGCACCAGGCTCGACGGCTCCGAGCTC 10647

Qy        721 GACGCGGCGTACTGGTACCGCAACATGTCGAGCACGGTCCGGTTCGAGCCCGCCGCCCGG 780
              |||| |||||||||| || ||| |   |  || |||| ||||  |  |||  ||  |
Db      10648 GACGGCGCGTACTGGTATCGAAACCTCCGGCAAACCGTCCTGTTCTCGAGCGCGACCGAG 10707

Qy        781 CTGCTTCTGCAGCAGGGGCCCAAGACGTTCGTCGAGATGAGCCCGCACCCGGTGCTGACC 840
              | ||| ||  |  | |||         | ||||||| | ||||| || || ||||| ||
Db      10708 CGGCTCCTCGACGATGGGCATCGCTTCTCCGTCGAGGTCAGCCCCCATCCCGTGCTCACG 10767

Qy        841 ATGGGCCTCCAGGAGCTCGCCCCGGACCTGGGCGACACCACCGGCACCGCCGACACCGT 899
              | | |||||  |||  |  |  | |          |   ||| |  || || | || |
Db      10768 CTCGCCCTCCGCGAGACCTGCGAGCGCTCACCGCTCGATCCCGTCGTCGTCGGCTCCAT 10826


RESULT 2
US-08-456-837-6
; Sequence 6, Application US/08456837
; Patent No. 5643774
;  GENERAL INFORMATION:
;    APPLICANT:  Schupp, Thomas
;    APPLICANT:  Ligon, James M.
;    APPLICANT:  Beck, James Joseph
;    APPLICANT:  Hill, Dwight Steven
;    APPLICANT:  Ryals, John Andrew
;    APPLICANT:  Gaffney, Thomas Deane
;    APPLICANT:  Lam, Stephen Ting
;    APPLICANT:  Hammer, Phillip E.
;    APPLICANT:  Uknes, Scott Joseph
;    TITLE OF INVENTION:  Genes for the synthesis of
;    TITLE OF INVENTION:  antipathogenic substances
;    NUMBER OF SEQUENCES:  22
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Ciba-Geigy Corporation
;      STREET:  7 Skyline Drive
;      CITY:  Hawthorne
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10532
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/456,837
;      FILING DATE:  01-JUN-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/457,205
;      FILING DATE:  01-JUN-1995
;      APPLICATION NUMBER:  08/258,261
;      FILING DATE:  08-Jun-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Elmer, James Scott
;      REGISTRATION NUMBER:  36,129
;      REFERENCE/DOCKET NUMBER:  CGC 1506/CIP3
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-541-8614
;      TELEFAX:  919-541-8689
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  28958 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
US-08-456-837-6

  Query Match            36.7%;  Score 348.2;  DB 1;  Length 28958;
  Best Local Similarity  62.7%;  Pred. No. 4.5e-57;
  Matches  564;  Conservative    0;  Mismatches  323;  Indels   12;  Gaps    1;

Qy          1 GTCTTCGTTTTTCCCGGGCAGGGCCCGCAATGGCCGGGCATGGGAAGGGAACTTCTCGAC 60
              |||||||| ||||| ||||| ||| |||| |||   || ||||         || ||||||
Db       9940 GTCTTCGTCTTTCCTGGGCAAGGCTCGCAGTGGGAAGGGATGGCCCTCTCCCTGCTCGAC 9999

Qy         61 GCTTCCGACGTCTTCCGGGAGAGCGTCCGCGCCTGCGAAGCCGCGTTCGCGCCCTACGTC 120
              | ||    |||||||| |       ||    || |||||   |||| |||| ||  |||||
Db      10000 TCCTCGCCGGTCTTCCGCGCTCAGCTCGAAGCATGCGAGCGCGCGCTCGCTCCTCACGTC 10059

Qy        121 GACTGGTCGGTGGAGCAGGTGTTGCGGGACTCGCCGGACGCTCCCGGGCTGGACCGGGTG 180
              || |||    ||       ||  ||||   |     || ||| |||   || ||||| ||
Db      10060 GAGTGGAGCCTGCTCGCCGTCCTGCGCCGCGACGAGGGCGCCCCCTCCCTCGACCGCGTC 10119

Qy        181 GACGTCGTCCAGCCGACCCTGTTCGCCGTCATGATCTCCCTGGCCGCCCTCTGGCGCTCG 240
              |||||||| |||||  |||| || ||||||||| ||||||||||||||||||||||||||
Db      10120 GACGTCGTACAGCCCGCCCTCTTTGCCGTCATGGTCTCCCTGGCCGCCCTCTGGCGCTCG 10179

Qy        241 CAAGGGGTCGAGCCGTGCGCGGTGCTGGGACACAGCCTGGGCGAGATCGCGGCAGCCCAC 300
```

```
Db      10180 CTCGGCGTCGAGCCCGCCGCCGTCGTCGGCCACAGCCAGGGCGAGATCGCCGCCGCCTTC 10239

Qy        301 GTCTCGGGAGGCCTGTCCCTGGCCGACGCCGCACGCGTGGTGACGCTTTGGAGCCAGGCA 360
              ||| | || |  || ||||| |  ||||| || ||| | |    | ||  | ||  | ||
Db      10240 GTCGCAGGCGCTCTCTCCCTCGAGGACGCGGCGCGCATCGCCGCCCTGCGCAGGAAAGCG 10299

Qy        361 CAGACCACCCTTGCCGGGACCGGCGCGCTCGTCTCCGTCGCCGCCACGCCGGATGAGCTC 420
              |  |||||| | | ||| | |||||    | | | ||||||     |     ||  |||
Db      10300 CTCACCACCGTCGGCGGCAACGGCGGCATGGCCGCCGTCGAGCTCGGCGCCTCCGACCTC 10359

Qy        421 CTGCCCCGAATCGCTCCGTGGACCGAGGACAACCCGGCGCGGCTCGCCGTCGCAGCCGTC 480
              | | ||    ||||||| |||  |||                 |||||  ||  ||| ||||||
Db      10360 CAGACCTACCTCGCTCCCTGGGGCGA------------CAGGCTCTCCACCGCCGCCGTC 10407

Qy        481 AACGGACCCCGGAGCACAGTCGTTTCCGGTGCCCGCGAGGCCGTCGCGGACCTGGTGGCC 540
              ||| | ||| ||   ||  |||| ||||| |  | ||  |||||||  |  ||| | | |
Db      10408 AACAGCCCCAGGGCTACCCTCGTATCCGGCGAGCCCGCCGCCGTCGACGCGCTGCTCGAC 10467

Qy        541 GACCTCACCGCCGCGCAGGTGCGCACGCGCATGATCCCGGTGGACGTTCCCGCCCACTCC 600
              | ||||||||||  |  |||||  | | |||| |||||  || |||     || ||||||||
Db      10468 GTCCTCACCGCCACCAAGGTGTTCGCCCGCAAGATCCGCGTCGACTACGCCTCCCACTCC 10527

Qy        601 CCCCTGATGTACGCCATCGAGGAACGGGTCGTCAGCGGCCTGCTGCCCATCACCCCACGC 660
               ||| |||| |||||  || | ||   | ||| |   || ||      |||| | || ||
Db      10528 GCCCAGATGGACGCCGTCCAAGACGAGCTCGCCGCAGGTCTAGCCAACATCGCTCCTCGG 10587

Qy        661 CCCTCCCGCATCCCCTTCCACTCCTCGGTGACCGGCGGCCGCCTCGACACCCGCGAGCTA 720
              | | |    ||||  |  | ||  | || ||||||  | |  |||||  |  ||||||
Db      10588 ACGTGCGAGCTCCCTCTTTATTCGACCGTCACCGGCACCAGGCTCGACGGCTCCGAGCTC 10647

Qy        721 GACGCGGCGTACTGGTACCGCAACATGTCGAGCACGGTCCGGTTCGAGCCCGCCGCCCGG 780
              ||||  |||||||||||| || ||| |   |   || |||| ||||  |  |||  ||  |
Db      10648 GACGGCGCGTACTGGTATCGAAACCTCCGGCAAACCGTCCTGTTCTCGAGCGCGACCGAG 10707

Qy        781 CTGCTTCTGCAGCAGGGGCCCAAGACGTTCGTCGAGATGAGCCCGCACCCGGTGCTGACC 840
              | ||| ||  |  | ||||         | |||||| | ||||| || || |||||  ||
Db      10708 CGGCTCCTCGACGATGGGCATCGCTTCTCCGTCGAGGTCAGCCCCCATCCCGTGCTCACG 10767

Qy        841 ATGGGCCTCCAGGAGCTCGCCCCGGACCTGGGCGACACCACCGGCACCGCCGACACCGT 899
               | | |||||  |||  |  |  |  |        |   ||| |  || || | || | |
Db      10768 CTCGCCCTCCGCGAGACCTGCGAGCGCTCACCGCTCGATCCCGTCGTCGTCGGCTCCAT 10826


RESULT 3
US-08-457-342-6
; Sequence 6, Application US/08457342
; Patent No. 5662898
;  GENERAL INFORMATION:
;    APPLICANT:  Schupp, Thomas
;    APPLICANT:  Ligon, James M.
;    APPLICANT:  Beck, James Joseph
;    APPLICANT:  Hill, Dwight Steven
;    APPLICANT:  Ryals, John Andrew
;    APPLICANT:  Gaffney, Thomas Deane
;    APPLICANT:  Lam, Stephen Ting
;    APPLICANT:  Hammer, Phillip E.
;    APPLICANT:  Uknes, Scott Joseph
;    TITLE OF INVENTION:  Genes for the synthesis of
;    TITLE OF INVENTION:  antipathogenic substances
;    NUMBER OF SEQUENCES:  22
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Ciba-Geigy Corporation
;      STREET:  7 Skyline Drive
;      CITY:  Hawthorne
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10532
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/457,342
;      FILING DATE:  01-JUN-1995
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/457,205
;      FILING DATE:  01-JUN-1995
;      APPLICATION NUMBER:  08/258,261
;      FILING DATE:  08-Jun-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Elmer, James Scott
;      REGISTRATION NUMBER:  36,129
;      REFERENCE/DOCKET NUMBER:  CGC 1506/CIP3
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-541-8614
;      TELEFAX:  919-541-8689
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  28958 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
US-08-457-342-6

  Query Match             36.7%;  Score 348.2;  DB 1;  Length 28958;
  Best Local Similarity   62.7%;  Pred. No. 4.5e-57;
  Matches  564;  Conservative    0;  Mismatches  323;  Indels   12;  Gaps    1;

Qy          1 GTCTTCGTTTTTCCCGGGCAGGGCCCGCAATGGCCGGGCATGGGAAGGGAACTTCTCGAC 60
              |||||||| ||||| ||||| ||| |||| |||    || ||||          || ||||||
Db       9940 GTCTTCGTCTTTCCTGGGCAAGGCTCGCAGTGGGAAGGGATGGCCCTCTCCCTGCTCGAC 9999

Qy         61 GCTTCCGACGTCTTCCGGGAGAGCGTCCGCGCCTGCGAAGCCGCGTTCGCGCCCTACGTC 120
               | ||   |||||||| |       ||   || |||||   |||| |||| ||  |||||
Db      10000 TCCTCGCCGGTCTTCCGCGCTCAGCTCGAAGCATGCGAGCGCGCGCTCGCTCCTCACGTC 10059

Qy        121 GACTGGTCGGTGGAGCAGGTGTTGCGGGACTCGCCGGACGCTCCCGGGCTGGACCGGGTG 180
              || |||    ||      ||  ||||    |      || ||| |||   || ||||| ||
Db      10060 GAGTGGAGCCTGCTCGCCGTCCTGCGCCGCGACGAGGGCGCCCCCTCCCTCGACCGCGTC 10119

Qy        181 GACGTCGTCCAGCCGACCCTGTTCGCCGTCATGATCTCCCTGGCCGCCCTCTGGCGCTCG 240
              |||||||| |||||   |||| || |||||||| ||||||||||||||||||||||||||||
Db      10120 GACGTCGTACAGCCCGCCCTCTTTGCCGTCATGGTCTCCCTGGCCGCCCTCTGGCGCTCG 10179

Qy        241 CAAGGGGTCGAGCCGTGCGCGGTGCTGGGACACAGCCTGGGCGAGATCGCGGCAGCCCAC 300
              |  || ||||||||   ||| ||  | || |||||| ||||||||||| || |||  |
Db      10180 CTCGGCGTCGAGCCCGCCGCCGTCGTCGGCCACAGCCAGGGCGAGATCGCCGCCGCCTTC 10239

Qy        301 GTCTCGGGAGGCCTGTCCCTGGCCGACGCCGCACGCGTGGTGACGCTTTGGAGCCAGGCA 360
              ||| | || |  || ||||| |  ||||| || ||| | |    | ||  | ||  | ||
Db      10240 GTCGCAGGCGCTCTCTCCCTCGAGGACGCGGCGCGCATCGCCGCCCTGCGCAGGAAAGCG 10299

Qy        361 CAGACCACCCTTGCCGGGACCGGCGCGCTCGTCTCCGTCGCCGCCACGCCGGATGAGCTC 420
              |  |||||| | | ||| | |||||    | | | ||||||     |     ||  |||
Db      10300 CTCACCACCGTCGGCGGCAACGGCGGCATGGCCGCCGTCGAGCTCGGCGCCTCCGACCTC 10359

Qy        421 CTGCCCCGAATCGCTCCGTGGACCGAGGACAACCCGGCGCGGCTCGCCGTCGCAGCCGTC 480
              | | ||    ||||||| |||  |||                 |||||  ||  ||| ||||||
Db      10360 CAGACCTACCTCGCTCCCTGGGGCGA------------CAGGCTCTCCACCGCCGCCGTC 10407

Qy        481 AACGGACCCCGGAGCACAGTCGTTTCCGGTGCCCGCGAGGCCGTCGCGGACCTGGTGGCC 540
              ||| | ||| ||   ||  |||| ||||| |  | ||  |||||||  |  ||| | | |
Db      10408 AACAGCCCCAGGGCTACCCTCGTATCCGGCGAGCCCGCCGCCGTCGACGCGCTGCTCGAC 10467

Qy        541 GACCTCACCGCCGCGCAGGTGCGCACGCGCATGATCCCGGTGGACGTTCCCGCCCACTCC 600
              | ||||||||||  |  |||||  | | |||| |||||  || |||     || ||||||||
Db      10468 GTCCTCACCGCCACCAAGGTGTTCGCCCGCAAGATCCGCGTCGACTACGCCTCCCACTCC 10527

Qy        601 CCCCTGATGTACGCCATCGAGGAACGGGTCGTCAGCGGCCTGCTGCCCATCACCCCACGC 660
               ||| |||| |||||  || | ||   | ||| |   || ||      |||| | || ||
```

```
Db      10528 GCCCAGATGGACGCCGTCCAAGACGAGCTCGCCGCAGGTCTAGCCAACATCGCTCCTCGG 10587

Qy        661 CCCTCCCGCATCCCCTTCCACTCCTCGGTGACCGGCGGCCGCCTCGACACCCGCGAGCTA 720

Db      10588 ACGTGCGAGCTCCCTCTTTATTCGACCGTCACCGGCACCAGGCTCGACGGCTCCGAGCTC 10647

Qy        721 GACGCGGCGTACTGGTACCGCAACATGTCGAGCACGGTCCGGTTCGAGCCCGCCGCCCGG 780

Db      10648 GACGGCGCGTACTGGTATCGAAACCTCCGGCAAACCGTCCTGTTCTCGAGCGCGACCGAG 10707

Qy        781 CTGCTTCTGCAGCAGGGGCCCAAGACGTTCGTCGAGATGAGCCCGCACCCGGTGCTGACC 840

Db      10708 CGGCTCCTCGACGATGGGCATCGCTTCTCCGTCGAGGTCAGCCCCCATCCCGTGCTCACG 10767

Qy        841 ATGGGCCTCCAGGAGCTCGCCCCGGACCTGGGCGACACCACCGGCACCGCCGACACCGT 899

Db      10768 CTCGCCCTCCGCGAGACCTGCGAGCGCTCACCGCTCGATCCCGTCGTCGTCGGCTCCAT 10826


RESULT 4
US-08-457-646A-6
; Sequence 6, Application US/08457646A
; Patent No. 5679560
;  GENERAL INFORMATION:
;    APPLICANT:  Schupp, Thomas
;    APPLICANT:  Ligon, James M.
;    APPLICANT:  Beck, James Joseph
;    APPLICANT:  Hill, Dwight Steven
;    APPLICANT:  Ryals, John Andrew
;    APPLICANT:  Gaffney, Thomas Deane
;    APPLICANT:  Lam, Stephen Ting
;    APPLICANT:  Hammer, Phillip E.
;    APPLICANT:  Uknes, Scott Joseph
;    TITLE OF INVENTION:  Genes for the synthesis of
;    TITLE OF INVENTION:  antipathogenic substances
;    NUMBER OF SEQUENCES:  22
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Ciba-Geigy Corporation
;      STREET:  7 Skyline Drive
;      CITY:  Hawthorne
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10532
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/457,646A
;      FILING DATE:  01-JUN-1995
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/457,205
;      FILING DATE:  01-JUN-1995
;      APPLICATION NUMBER:  08/258,261
;      FILING DATE:  08-Jun-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Elmer, James Scott
;      REGISTRATION NUMBER:  36,129
;      REFERENCE/DOCKET NUMBER:  CGC 1506/CIP3
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-541-8614
;      TELEFAX:  919-541-8689
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  28958 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO

US-08-457-646A-6

  Query Match           36.7%;  Score 348.2;  DB 1;  Length 28958;
  Best Local Similarity  62.7%;  Pred. No. 4.5e-57;
  Matches  564;  Conservative    0;  Mismatches  323;  Indels   12;  Gaps    1;

Qy          1 GTCTTCGTTTTTCCCGGGCAGGGCCCGCAATGGCCGGGCATGGGAAGGGAACTTCTCGAC 60

Db       9940 GTCTTCGTCTTTCCTGGGCAAGGCTCGCAGTGGGAAGGGATGGCCCTCTCCCTGCTCGAC 9999

Qy         61 GCTTCCGACGTCTTCCGGGAGAGCGTCCGCGCCTGCGAAGCCGCGTTCGCGCCCTACGTC 120

Db      10000 TCCTCGCCGGTCTTCCGCGCTCAGCTCGAAGCATGCGAGCGCGCGCTCGCTCCTCACGTC 10059

Qy         121 GACTGGTCGGTGGAGCAGGTGTTGCGGGACTCGCCGGACGCTCCCGGGCTGGACCGGGTG 180

Db      10060 GAGTGGAGCCTGCTCGCCGTCCTGCGCCGCGACGAGGGCGCCCCCTCCCTCGACCGCGTC 10119

Qy         181 GACGTCGTCCAGCCGACCCTGTTCGCCGTCATGATCTCCCTGGCCGCCCTCTGGCGCTCG 240

Db      10120 GACGTCGTACAGCCCGCCCTCTTTGCCGTCATGGTCTCCCTGGCCGCCCTCTGGCGCTCG 10179

Qy         241 CAAGGGGTCGAGCCGTGCGCGGTGCTGGGACACAGCCTGGGCGAGATCGCGGCAGCCCAC 300

Db      10180 CTCGGCGTCGAGCCCGCCGCCGTCGTCGGCCACAGCCAGGGCGAGATCGCCGCCGCCTTC 10239

Qy         301 GTCTCGGGAGGCCTGTCCCTGGCCGACGCCGCACGCGTGGTGACGCTTTGGAGCCAGGCA 360

Db      10240 GTCGCAGGCGCTCTCTCCCTCGAGGACGCGGCGCGCATCGCCGCCCTGCGCAGGAAAGCG 10299

Qy         361 CAGACCACCCTTGCCGGGACCGGCGCGCTCGTCTCCGTCGCCGCCACGCCGGATGAGCTC 420

Db      10300 CTCACCACCGTCGGCGGCAACGGCGGCATGGCCGCCGTCGAGCTCGGCGCCTCCGACCTC 10359

Qy         421 CTGCCCCGAATCGCTCCGTGGACCGAGGACAACCCGGCGCGGCTCGCCGTCGCAGCCGTC 480

Db      10360 CAGACCTACCTCGCTCCCTGGGGCGA-----------CAGGCTCTCCACCGCCGCCGTC 10407

Qy         481 AACGGACCCCGGAGCACAGTCGTTTCCGGTGCCCGCGAGGCCGTCGCGGACCTGGTGGCC 540

Db      10408 AACAGCCCCAGGGCTACCCTCGTATCCGGCGAGCCCGCCGCCGTCGACGCGCTGCTCGAC 10467

Qy         541 GACCTCACCGCCGCGCAGGTGCGCACGCGCATGATCCCGGTGGACGTTCCCGCCCACTCC 600

Db      10468 GTCCTCACCGCCACCAAGGTGTTCGCCCGCAAGATCCGCGTCGACTACGCCTCCCACTCC 10527

Qy         601 CCCCTGATGTACGCCATCGAGGAACGGGTCGTCAGCGGCCTGCTGCCCCATCACCCCACGC 660

Db      10528 GCCCAGATGGACGCCGTCCAAGACGAGCTCGCCGCAGGTCTAGCCAACATCGCTCCTCGG 10587

Qy         661 CCCTCCCGCATCCCCTTCCACTCCTCGGTGACCGGCGGCCGCCTCGACACCCGCGAGCTA 720

Db      10588 ACGTGCGAGCTCCCTCTTTATTCGACCGTCACCGGCACCAGGCTCGACGGCTCCGAGCTC 10647

Qy         721 GACGCGGCGTACTGGTACCGCAACATGTCGAGCACGGTCCGGTTCGAGCCCGCCGCCCGG 780

Db      10648 GACGGCGCGTACTGGTATCGAAACCTCCGGCAAACCGTCCTGTTCTCGAGCGCGACCGAG 10707

Qy         781 CTGCTTCTGCAGCAGGGGCCCAAGACGTTCGTCGAGATGAGCCCGCACCCGGTGCTGACC 840

Db      10708 CGGCTCCTCGACGATGGGCATCGCTTCTCCGTCGAGGTCAGCCCCCATCCCGTGCTCACG 10767

Qy         841 ATGGGCCTCCAGGAGCTCGCCCCGGACCTGGGCGACACCACCGGCACCGCCGACACCGT 899

Db      10768 CTCGCCCTCCGCGAGACCTGCGAGCGCTCACCGCTCGATCCCGTCGTCGTCGGCTCCAT 10826


RESULT 5
US-08-458-076A-6
; Sequence 6, Application US/08458076A
; Patent No. 5698425
;  GENERAL INFORMATION:
;    APPLICANT:  Schupp, Thomas
```

```
;    APPLICANT:  Ligon, James M.
;    APPLICANT:  Beck, James Joseph
;    APPLICANT:  Hill, Dwight Steven
;    APPLICANT:  Ryals, John Andrew
;    APPLICANT:  Gaffney, Thomas Deane
;    APPLICANT:  Lam, Stephen Ting
;    APPLICANT:  Hammer, Phillip E.
;    APPLICANT:  Uknes, Scott Joseph
;    TITLE OF INVENTION:  Genes for the synthesis of
;    TITLE OF INVENTION:  antipathogenic substances
;    NUMBER OF SEQUENCES:  22
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Ciba-Geigy Corporation
;      STREET:  7 Skyline Drive
;      CITY:  Hawthorne
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10532
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/458,076A
;      FILING DATE:  01-JUN-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/457,205
;      FILING DATE:  01-JUN-1995
;      APPLICATION NUMBER:  08/258,261
;      FILING DATE:  08-Jun-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Elmer, James Scott
;      REGISTRATION NUMBER:  36,129
;      REFERENCE/DOCKET NUMBER:  CGC 1506/CIP3
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-541-8614
;      TELEFAX:  919-541-8689
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  28958 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
US-08-458-076A-6

  Query Match             36.7%;  Score 348.2;  DB 1;  Length 28958;
  Best Local Similarity   62.7%;  Pred. No. 4.5e-57;
  Matches  564;  Conservative    0;  Mismatches  323;  Indels   12;  Gaps    1;

Qy          1 GTCTTCGTTTTTCCCGGGCAGGGCCCGCAATGGCCGGGCATGGGAAGGGAACTTCTCGAC 60
              |||||||| ||||| ||||| ||| |||| |||   || ||||         || ||||||
Db       9940 GTCTTCGTCTTTCCTGGGCAAGGCTCGCAGTGGGAAGGGATGGCCCTCTCCCTGCTCGAC 9999

Qy         61 GCTTCCGACGTCTTCCGGGAGAGCGTCCGCGCCTGCGAAGCCGCGTTCGCGCCCTACGTC 120
                | ||    ||||||||  |      ||   || |||||   |||| |||| ||  |||||
Db      10000 TCCTCGCCGGTCTTCCGCGCTCAGCTCGAAGCATGCGAGCGCGCGCTCGCTCCTCACGTC 10059

Qy        121 GACTGGTCGGTGGAGCAGGTGTTGCGGGACTCGCCGGACGCTCCCGGGCTGGACCGGGTG 180
              || |||    ||      ||  ||||   |     || ||| |||   || |||||  ||
Db      10060 GAGTGGAGCCTGCTCGCCGTCCTGCGCCGCGACGAGGGCGCCCCCTCCCTCGACCGCGTC 10119

Qy        181 GACGTCGTCCAGCCGACCCTGTTCGCCGTCATGATCTCCCTGGCCGCCCTCTGGCGCTCG 240
              |||||||| |||||  |||| || |||||||||| ||||||||||||||||||||||||||
Db      10120 GACGTCGTACAGCCCGCCCTCTTTGCCGTCATGGTCTCCCTGGCCGCCCTCTGGCGCTCG 10179

Qy        241 CAAGGGGTCGAGCCGTGCGCGGTGCTGGGACACAGCCTGGGCGAGATCGCGGCAGCCCAC 300
              |  || ||||||||   ||| ||  | || ||||||| |||||||||||| || |||  |
Db      10180 CTCGGCGTCGAGCCCGCCGCCGTCGTCGGCCACAGCCAGGGCGAGATCGCCGCCGCCTTC 10239

Qy        301 GTCTCGGGAGGCCTGTCCCTGGCCGACGCCGCACGCGTGGTGACGCTTTGGAGCCAGGCA 360
              ||| | || |  || ||||| |  ||||| || ||| | |    | ||  | ||  | ||
Db      10240 GTCGCAGGCGCTCTCTCCCTCGAGGACGCGGCGCGCATCGCCGCCCTGCGCAGGAAAGCG 10299

Qy        361 CAGACCACCCTTGCCGGGACCGGCGCGCTCGTCTCCGTCGCCGCCACGCCGGATGAGCTC 420
              |  |||||| | | ||| | |||||    | | | ||||||     |     |   || |||
Db      10300 CTCACCACCGTCGGCGGCAACGGCGGCATGGCCGCCGTCGAGCTCGGCGCCTCCGACCTC 10359

Qy        421 CTGCCCCGAATCGCTCCGTGGACCGAGGACAACCCGGCGCGGCTCGCCGTCGCAGCCGTC 480
              | | ||    ||||||| |||  |||                ||||| ||  ||| ||||||
Db      10360 CAGACCTACCTCGCTCCCTGGGGCGA-----------CAGGCTCTCCACCGCCGCCGTC 10407

Qy        481 AACGGACCCCGGAGCACAGTCGTTTCCGGTGCCCGCGAGGCCGTCGCGGACCTGGTGGCC 540
              ||| | ||| ||   ||  |||| |||||  |  | ||  |||||||  |  ||| | | |
Db      10408 AACAGCCCCAGGGCTACCCTCGTATCCGGCGAGCCCGCCGCCGTCGACGCGCTGCTCGAC 10467

Qy        541 GACCTCACCGCCGCGCAGGTGCGCACGCGCATGATCCCGGTGGACGTTCCCGCCCACTCC 600
              | |||||||||| |  |||||  | | |||| |||||  || |||     || ||||||||
Db      10468 GTCCTCACCGCCACCAAGGTGTTCGCCCGCAAGATCCGCGTCGACTACGCCTCCCACTCC 10527

Qy        601 CCCCTGATGTACGCCATCGAGGAACGGGTCGTCAGCGGCCTGCTGCCCATCACCCCACGC 660
              ||| |||| ||||| || | ||   | ||| |   || ||       |||| | || ||
Db      10528 GCCCAGATGGACGCCGTCCAAGACGAGCTCGCCGCAGGTCTAGCCAACATCGCTCCTCGG 10587

Qy        661 CCCTCCCGCATCCCCTTCCACTCCTCGGTGACCGGCGGCCGCCTCGACACCCGCGAGCTA 720
              | | |    ||||  |  | ||  | || ||||||  | | ||||||  |  ||||||
Db      10588 ACGTGCGAGCTCCCTCTTTATTCGACCGTCACCGGCACCAGGCTCGACGGCTCCGAGCTC 10647

Qy        721 GACGCGGCGTACTGGTACCGCAACATGTCGAGCACGGTCCGGTTCGAGCCCGCCGCCCGG 780
              ||||  |||||||||||| || ||| |   |    || |||| |||| |  ||| ||  |
Db      10648 GACGGCGCGTACTGGTATCGAAACCTCCGGCAAACCGTCCTGTTCTCGAGCGCGACCGAG 10707

Qy        781 CTGCTTCTGCAGCAGGGGCCCAAGACGTTCGTCGAGATGAGCCCGCACCCGGTGCTGACC 840
              | ||| ||  |  | ||||        | ||||||| | ||||| || || ||||| ||
Db      10708 CGGCTCCTCGACGATGGGCATCGCTTCTCCGTCGAGGTCAGCCCCCATCCCGTGCTCACG 10767

Qy        841 ATGGGCCTCCAGGAGCTCGCCCCGGACCTGGGCGACACCACCGGCACCGCCGACACCGT 899
              | || |||||  |||  |  |  |  |         |    ||| |  || || | || |
Db      10768 CTCGCCCTCCGCGAGACCTGCGAGCGCTCACCGCTCGATCCCGTCGTCGTCGGCTCCAT 10826


RESULT 6
US-08-764-233A-4
; Sequence 4, Application US/08764233A
; Patent No. 5716849
;  GENERAL INFORMATION:
;    APPLICANT:  Ligon, James M.
;    APPLICANT:  Schupp, Thomas
;    APPLICANT:  Beck, James J.
;    APPLICANT:  Hill, Dwight S.
;    APPLICANT:  Neff, Snezanna
;    APPLICANT:  Ryals, John A.
;    TITLE OF INVENTION:  Genes For The Biosynthesis Of Soraphen
;    NUMBER OF SEQUENCES:  10
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Ciba-Geigy Corporation
;      STREET:  520 White Plains Road, P.O. Box 2005
;      CITY:  Tarrytown
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10591
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/764,233A
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/729,214
```

```
;      FILING DATE:  09-OCT-1996
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/258,261
;      FILING DATE:  08-JUN-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Meigs, J. Timothy
;      REGISTRATION NUMBER:  38,241
;      REFERENCE/DOCKET NUMBER:  1506/CIP6
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (919) 541-8587
;      TELEFAX:  (919) 541-8689
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  28958 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    ORIGINAL SOURCE:
;      ORGANISM:  Sorangium cellulosum
;    IMMEDIATE SOURCE:
;      CLONE:  p98/1
US-08-764-233A-4

  Query Match           36.7%;  Score 348.2;  DB 1;  Length 28958;
  Best Local Similarity   62.7%;  Pred. No. 4.5e-57;
  Matches  564;  Conservative    0;  Mismatches  323;  Indels   12;  Gaps    1;

Qy          1 GTCTTCGTTTTTCCCGGGCAGGGCCCGCAATGGCCGGGCATGGGAAGGGAACTTCTCGAC 60
              |||||||| ||||| ||||| ||| |||| |||    || ||||         || ||||||
Db       9940 GTCTTCGTCTTTCCTGGGCAAGGCTCGCAGTGGGAAGGGATGGCCCTCTCCCTGCTCGAC 9999

Qy         61 GCTTCCGACGTCTTCCGGGAGAGCGTCCGCGCCTGCGAAGCCGCGTTCGCGCCCTACGTC 120
               | ||    |||||||| |       ||   || |||||   |||| |||| ||  |||||
Db      10000 TCCTCGCCGGTCTTCCGCGCTCAGCTCGAAGCATGCGAGCGCGCGCTCGCTCCTCACGTC 10059

Qy        121 GACTGGTCGGTGGAGCAGGTGTTGCGGGACTCGCCGGACGCTCCCGGGCTGGACCGGGTG 180
              || |||    ||       ||  ||||   |     || ||| |||   || ||||| ||
Db      10060 GAGTGGAGCCTGCTCGCCGTCCTGCGCCGCGACGAGGGCGCCCCCTCCCTCGACCGCGTC 10119

Qy        181 GACGTCGTCCAGCCGACCCTGTTCGCCGTCATGATCTCCCTGGCCGCCCTCTGGCGCTCG 240
              |||||||| |||||  |||| || ||||||||| ||||||||||||||||||||||||||
Db      10120 GACGTCGTACAGCCCGCCCTCTTTGCCGTCATGGTCTCCCTGGCCGCCCTCTGGCGCTCG 10179

Qy        241 CAAGGGGTCGAGCCGTGCGCGGTGCTGGGACACAGCCTGGGCGAGATCGCGGCAGCCCAC 300
              |  || ||||||||   ||| ||  | || ||||||| |||||||||||| || |||  |
Db      10180 CTCGGCGTCGAGCCCGCCGCCGTCGTCGGCCACAGCCAGGGCGAGATCGCCGCCGCCTTC 10239

Qy        301 GTCTCGGGAGGCCTGTCCCTGGCCGACGCCGCACGCGTGGTGACGCTTTGGAGCCAGGCA 360
              ||| | || |  || ||||| |  ||||| || ||| | |   | ||  | ||  | ||
Db      10240 GTCGCAGGCGCTCTCTCCCTCGAGGACGCGGCGCGCATCGCCGCCCTGCGCAGGAAAGCG 10299

Qy        361 CAGACCACCCTTGCCGGGACCGGCGCGCTCGTCTCCGTCGCCGCCACGCCGGATGAGCTC 420
              |  |||||| | | ||| | |||||   | | | ||||||    |     |    || |||
Db      10300 CTCACCACCGTCGGCGGCAACGGCGGCATGGCCGCCGTCGAGCTCGGCGCCTCCGACCTC 10359

Qy        421 CTGCCCCGAATCGCTCCGTGGACCGAGGACAACCCGGCGCGGCTCGCCGTCGCAGCCGTC 480
              | | ||   |||||| |||  |||              ||||| ||  ||| ||||||
Db      10360 CAGACCTACCTCGCTCCCTGGGGCGA----------CAGGCTCTCCACCGCCGCCGTC 10407

Qy        481 AACGGACCCCGGAGCACAGTCGTTTCCGGTGCCCGCGAGGCCGTCGCGGACCTGGTGGCC 540
              ||| | ||| ||    ||  |||| ||||| |  | ||  |||||||  |  ||| | | |
Db      10408 AACAGCCCCAGGGCTACCCTCGTATCCGGCGAGCCCGCCGCCGTCGACGCGCTGCTCGAC 10467

Qy        541 GACCTCACCGCCGCGCAGGTGCGCACGCGCATGATCCCGGTGGACGTTCCCGCCCACTCC 600
              | |||||||||| |  |||||  | | |||| ||||  || |||    || ||||||||
Db      10468 GTCCTCACCGCCACCAAGGTGTTCGCCCGCAAGATCCGCGTCGACTACGCCTCCCACTCC 10527

Qy        601 CCCCTGATGTACGCCATCGAGGAACGGGTCGTCAGCGGCCTGCTGCCCATCACCCCACGC 660
               ||| |||| ||||| || | ||   | ||| |   || ||      |||| | || ||
Db      10528 GCCCAGATGGACGCCGTCCAAGACGAGCTCGCCGCAGGTCTAGCCAACATCGCTCCTCGG 10587

Qy        661 CCCTCCCGCATCCCCTTCCACTCCTCGGTGACCGGCGGCCGCCTCGACACCCGCGAGCTA 720
              | | |   ||||  |  | ||  | || ||||||  | | ||||||  |  ||||||
Db      10588 ACGTGCGAGCTCCCTCTTTATTCGACCGTCACCGGCACCAGGCTCGACGGCTCCGAGCTC 10647

Qy        721 GACGCGGCGTACTGGTACCGCAACATGTCGAGCACGGTCCGGTTCGAGCCCGCCGCCCGG 780
              ||||  |||||||||||| || ||| |   |    || |||| ||||  |  |||  || |
Db      10648 GACGGCGCGTACTGGTATCGAAACCTCCGGCAAACCGTCCTGTTCTCGAGCGCGACCGAG 10707

Qy        781 CTGCTTCTGCAGCAGGGGCCCAAGACGTTCGTCGAGATGAGCCCGCACCCGGTGCTGACC 840
              | ||| ||  |  | ||||      | |||||| | ||||| || || ||||| ||
Db      10708 CGGCTCCTCGACGATGGGCATCGCTTCTCCGTCGAGGTCAGCCCCCATCCCGTGCTCACG 10767

Qy        841 ATGGGCCTCCAGGAGCTCGCCCCGGACCTGGGCGACACCACCGGCACCGCCGACACCGT 899
              | | |||||  |||  |  |  |  |        |    ||| |  || || | || |
Db      10768 CTCGCCCTCCGCGAGACCTGCGAGCGCTCACCGCTCGATCCCGTCGTCGTCGGCTCCAT 10826


RESULT 7
US-08-457-335A-6
; Sequence 6, Application US/08457335A
; Patent No. 5723759
;  GENERAL INFORMATION:
;    APPLICANT:  Schupp, Thomas
;    APPLICANT:  Ligon, James M.
;    APPLICANT:  Beck, James Joseph
;    APPLICANT:  Hill, Dwight Steven
;    APPLICANT:  Ryals, John Andrew
;    APPLICANT:  Gaffney, Thomas Deane
;    APPLICANT:  Lam, Stephen Ting
;    APPLICANT:  Hammer, Phillip E.
;    APPLICANT:  Uknes, Scott Joseph
;    TITLE OF INVENTION:  Genes for the synthesis of
;    TITLE OF INVENTION:  antipathogenic substances
;    NUMBER OF SEQUENCES:  22
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Ciba-Geigy Corporation
;      STREET:  7 Skyline Drive
;      CITY:  Hawthorne
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10532
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/457,335A
;      FILING DATE:  01-JUN-1995
;      CLASSIFICATION:  800
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/457,205
;      FILING DATE:  01-JUN-1995
;      APPLICATION NUMBER:  08/258,261
;      FILING DATE:  08-Jun-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Elmer, James Scott
;      REGISTRATION NUMBER:  36,129
;      REFERENCE/DOCKET NUMBER:  CGC 1506/CIP3
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-541-8614
;      TELEFAX:  919-541-8689
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  28958 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
US-08-457-335A-6

  Query Match           36.7%;  Score 348.2;  DB 1;  Length 28958;
```

```
  Best Local Similarity   62.7%;  Pred. No. 4.5e-57;
  Matches  564;  Conservative    0;  Mismatches  323;  Indels   12;  Gaps    1;

Qy          1 GTCTTCGTTTTTCCCGGGCAGGGCCCGCAATGGCCGGGCATGGGAAGGGAACTTCTCGAC 60
              |||||||| ||||| |||| |||| ||| ||| ||| |||| |||| |||| ||||||
Db       9940 GTCTTCGTCTTTCCTGGGCAAGGCTCGCAGTGGGAAGGGATGGCCCTCTCCCTGCTCGAC 9999

Qy         61 GCTTCCGACGTCTTCCGGGAGAGCGTCCGCGCCTGCGAAGCCGCGTTCGCGCCCTACGTC 120
                | ||    |||||||| |       ||   || |||||   |||| |||| ||  |||||
Db      10000 TCCTCGCCGGTCTTCCGCGCTCAGCTCGAAGCATGCGAGCGCGCGCTCGCTCCTCACGTC 10059

Qy        121 GACTGGTCGGTGGAGCAGGTGTTGCGGGACTCGCCGGACGCTCCCGGGCTGGACCGGGTG 180
              || |||     ||       ||  ||||   |      || ||| |||   || |||||  ||
Db      10060 GAGTGGAGCCTGCTCGCCGTCCTGCGCCGCGACGAGGGCGCCCCCTCCCTCGACCGCGTC 10119

Qy        181 GACGTCGTCCAGCCGACCCTGTTCGCCGTCATGATCTCCCTGGCCGCCCTCTGGCGCTCG 240
              |||||||| |||||  |||| || |||||||| ||||||||||||||||||||||||||||
Db      10120 GACGTCGTACAGCCCGCCCTCTTTGCCGTCATGGTCTCCCTGGCCGCCCTCTGGCGCTCG 10179

Qy        241 CAAGGGGTCGAGCCGTGCGCGGTGCTGGGACACAGCCTGGGCGAGATCGCGGCAGCCCAC 300
              |  || ||||||||   ||| ||  | || ||||||| |||||||||||| || |||  |
Db      10180 CTCGGCGTCGAGCCCGCCGCCGTCGTCGGCCACAGCCAGGGCGAGATCGCCGCCGCCTTC 10239

Qy        301 GTCTCGGGAGGCCTGTCCCTGGCCGACGCCGCACGCGTGGTGACGCTTTGGAGCCAGGCA 360
              ||| | || |  || ||||| |  ||||| || ||| | |   | ||  | ||  | ||
Db      10240 GTCGCAGGCGCTCTCTCCCTCGAGGACGCGGCGCGCATCGCCGCCCTGCGCAGGAAAGCG 10299

Qy        361 CAGACCACCCTTGCCGGGACCGGCGCGCTCGTCTCCGTCGCCGCCACGCCGGATGAGCTC 420
              |  |||||| | | ||| | |||||    | | | ||||||     |     |   || |||
Db      10300 CTCACCACCGTCGGCGGCAACGGCGGCATGGCCGCCGTCGAGCTCGGCGCCTCCGACCTC 10359

Qy        421 CTGCCCCGAATCGCTCCGTGGACCGAGGACAACCCGGCGCGGCTCGCCGTCGCAGCCGTC 480
              | | ||    ||||||| |||  |||                |||||  ||  ||| ||||||
Db      10360 CAGACCTACCTCGCTCCCTGGGGCGA----------CAGGCTCTCCACCGCCGCCGTC 10407

Qy        481 AACGGACCCCGGAGCACAGTCGTTTCCGGTGCCCGCGAGGCCGTCGCGGACCTGGTGGCC 540
              ||| | ||| ||   ||  |||| |||||  | | ||  |||||||  |  ||| | | |
Db      10408 AACAGCCCCAGGGCTACCCTCGTATCCGGCGAGCCCGCCGCCGTCGACGCGCTGCTCGAC 10467

Qy        541 GACCTCACCGCCGCGCAGGTGCGCACGCGCATGATCCCGGTGGACGTTCCCGCCCACTCC 600
              | ||||||||||  |  |||||  | | |||| |||||  || |||    || ||||||||
Db      10468 GTCCTCACCGCCACCAAGGTGTTCGCCCGCAAGATCCGCGTCGACTACGCCTCCCACTCC 10527

Qy        601 CCCCTGATGTACGCCATCGAGGAACGGGTCGTCAGCGGCCTGCTGCCCATCACCCCACGC 660
              |||  |||| ||||| || | ||   | |||  |  || ||      |||| | || ||
Db      10528 GCCCAGATGGACGCCGTCCAAGACGAGCTCGCCGCAGGTCTAGCCAACATCGCTCCTCGG 10587

Qy        661 CCCTCCCGCATCCCCTTCCACTCCTCGGTGACCGGCGGCCGCCTCGACACCCGCGAGCTA 720
              | | |   |||| |  | ||  | || ||||||  | | ||||||  |  ||||||
Db      10588 ACGTGCGAGCTCCCTCTTTATTCGACCGTCACCGGCACCAGGCTCGACGGCTCCGAGCTC 10647

Qy        721 GACGCGGCGTACTGGTACCGCAACATGTCGAGCACGGTCCGGTTCGAGCCCGCCGCCCGG 780
              ||||  |||||||||| || ||| |   |  || |||| |||| |  |||  ||  |
Db      10648 GACGGCGCGTACTGGTATCGAAACCTCCGGCAAACCGTCCTGTTCTCGAGCGCGACCGAG 10707

Qy        781 CTGCTTCTGCAGCAGGGGCCCAAGACGTTCGTCGAGATGAGCCCGCACCCGGTGCTGACC 840
              | ||| ||  |  | ||||       | ||||||| | ||||| || || ||||| ||
Db      10708 CGGCTCCTCGACGATGGGCATCGCTTCTCCGTCGAGGTCAGCCCCCATCCCGTGCTCACG 10767

Qy        841 ATGGGCCTCCAGGAGCTCGCCCCGGACCTGGGCGACACCACCGGCACCGCCGACACCGT 899
              | | |||||  |||  |  | | |  |       |   ||| |  || || | || |
Db      10768 CTCGCCCTCCGCGAGACCTGCGAGCGCTCACCGCTCGATCCCGTCGTCGTCGGCTCCAT 10826


RESULT 8
US-08-729-214-6
; Sequence 6, Application US/08729214
; Patent No. 5817502
;  GENERAL INFORMATION:
;    APPLICANT:  Ligon, James M.
;    APPLICANT:  Hill, Dwight Steven
;    APPLICANT:  Ryals, John Andrew
;    APPLICANT:  Hammer, Phillip E.
;    APPLICANT:  van Pee, Karl-Heinz
;    APPLICANT:  Kirner, Sabine
;    TITLE OF INVENTION:  Genes for the synthesis of
;    TITLE OF INVENTION:  antipathogenic substances
;    NUMBER OF SEQUENCES:  27
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Ciba-Geigy Corporation
;      STREET:  520 White Plains Road
;      CITY:  Tarrytown
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10591
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/729,214
;      FILING DATE:  TBA
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Meigs, J. Timothy
;      REGISTRATION NUMBER:  38,241
;      REFERENCE/DOCKET NUMBER:  CGC 1506/CIP5
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-541-8587
;      TELEFAX:  919-541-8689
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  28958 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
US-08-729-214-6

  Query Match             36.7%;  Score 348.2;  DB 1;  Length 28958;
  Best Local Similarity   62.7%;  Pred. No. 4.5e-57;
  Matches  564;  Conservative    0;  Mismatches  323;  Indels   12;  Gaps    1;

Qy          1 GTCTTCGTTTTTCCCGGGCAGGGCCCGCAATGGCCGGGCATGGGAAGGGAACTTCTCGAC 60
              |||||||| ||||| ||||| ||| |||| |||    || ||||         || ||||||
Db       9940 GTCTTCGTCTTTCCTGGGCAAGGCTCGCAGTGGGAAGGGATGGCCCTCTCCCTGCTCGAC 9999

Qy         61 GCTTCCGACGTCTTCCGGGAGAGCGTCCGCGCCTGCGAAGCCGCGTTCGCGCCCTACGTC 120
                | ||    |||||||| |       ||   || |||||   |||| |||| ||  |||||
Db      10000 TCCTCGCCGGTCTTCCGCGCTCAGCTCGAAGCATGCGAGCGCGCGCTCGCTCCTCACGTC 10059

Qy        121 GACTGGTCGGTGGAGCAGGTGTTGCGGGACTCGCCGGACGCTCCCGGGCTGGACCGGGTG 180
              || |||     ||       ||  ||||   |      || ||| |||   || |||||  ||
Db      10060 GAGTGGAGCCTGCTCGCCGTCCTGCGCCGCGACGAGGGCGCCCCCTCCCTCGACCGCGTC 10119

Qy        181 GACGTCGTCCAGCCGACCCTGTTCGCCGTCATGATCTCCCTGGCCGCCCTCTGGCGCTCG 240
              |||||||| |||||  |||| || |||||||| ||||||||||||||||||||||||||||
Db      10120 GACGTCGTACAGCCCGCCCTCTTTGCCGTCATGGTCTCCCTGGCCGCCCTCTGGCGCTCG 10179

Qy        241 CAAGGGGTCGAGCCGTGCGCGGTGCTGGGACACAGCCTGGGCGAGATCGCGGCAGCCCAC 300
              |  || ||||||||   ||| ||  | || ||||||| |||||||||||| || |||  |
Db      10180 CTCGGCGTCGAGCCCGCCGCCGTCGTCGGCCACAGCCAGGGCGAGATCGCCGCCGCCTTC 10239

Qy        301 GTCTCGGGAGGCCTGTCCCTGGCCGACGCCGCACGCGTGGTGACGCTTTGGAGCCAGGCA 360
              ||| | || |  || ||||| |  ||||| || ||| | |   | ||  | ||  | ||
Db      10240 GTCGCAGGCGCTCTCTCCCTCGAGGACGCGGCGCGCATCGCCGCCCTGCGCAGGAAAGCG 10299

Qy        361 CAGACCACCCTTGCCGGGACCGGCGCGCTCGTCTCCGTCGCCGCCACGCCGGATGAGCTC 420
              |  |||||| | | ||| | |||||    | | | ||||||     |     |   || |||
Db      10300 CTCACCACCGTCGGCGGCAACGGCGGCATGGCCGCCGTCGAGCTCGGCGCCTCCGACCTC 10359

Qy        421 CTGCCCCGAATCGCTCCGTGGACCGAGGACAACCCGGCGCGGCTCGCCGTCGCAGCCGTC 480
              | | ||    ||||||| |||  |||                |||||  ||  ||| ||||||
```

```
Db      10360 CAGACCTACCTCGCTCCCTGGGGCGA-----------CAGGCTCTCCACCGCCGCCGTC 10407

Qy        481 AACGGACCCCGGAGCACAGTCGTTTCCGGTGCCCGCGAGGCCGTCGCGGACCTGGTGGCC 540

Db      10408 AACAGCCCCAGGGCTACCCTCGTATCCGGCGAGCCCGCCGCCGTCGACGCGCTGCTCGAC 10467

Qy        541 GACCTCACCGCCGCGCAGGTGCGCACGCGCATGATCCCGGTGGACGTTCCCGCCCACTCC 600

Db      10468 GTCCTCACCGCCACCAAGGTGTTCGCCCGCAAGATCCGCGTCGACTACGCCTCCCACTCC 10527

Qy        601 CCCCTGATGTACGCCATCGAGGAACGGGTCGTCAGCGGCCTGCTGCCCATCACCCCACGC 660

Db      10528 GCCCAGATGGACGCCGTCCAAGACGAGCTCGCCGCAGGTCTAGCCAACATCGCTCCTCGG 10587

Qy        661 CCCTCCCGCATCCCCTTCCACTCCTCGGTGACCGGCGGCCGCCTCGACACCCGCGAGCTA 720

Db      10588 ACGTGCGAGCTCCCTCTTTATTCGACCGTCACCGGCACCAGGCTCGACGGCTCCGAGCTC 10647

Qy        721 GACGCGGCGTACTGGTACCGCAACATGTCGAGCACGGTCCGGTTCGAGCCCGCCGCCCGG 780

Db      10648 GACGGCGCGTACTGGTATCGAAACCTCCGGCAAACCGTCCTGTTCTCGAGCGCGACCGAG 10707

Qy        781 CTGCTTCTGCAGCAGGGGCCCAAGACGTTCGTCGAGATGAGCCCGCACCCGGTGCTGACC 840

Db      10708 CGGCTCCTCGACGATGGGCATCGCTTCTCCGTCGAGGTCAGCCCCCATCCCGTGCTCACG 10767

Qy        841 ATGGGCCTCCAGGAGCTCGCCCCGGACCTGGGCGACACCACCGGCACCGCCGACACCGT 899

Db      10768 CTCGCCCTCCGCGAGACCTGCGAGCGCTCACCGCTCGATCCCGTCGTCGTCGGCTCCAT 10826


RESULT 9
US-09-028-934-6
; Sequence 6, Application US/09028934
; Patent No. 6117670
;  GENERAL INFORMATION:
;    APPLICANT:  Ligon, James M.
;    APPLICANT:  Hill, Dwight S.
;    APPLICANT:  Lam, Steven T.
;    APPLICANT:  Hammer, Philip E.
;    APPLICANT:  van Pee, Karl-Heinz
;    APPLICANT:  Kirner, Sabine
;    APPLICANT:  Young, Thomas R.
;    TITLE OF INVENTION:  Pyrrolnitrin Biosynthesis Genes and Uses
;    TITLE OF INVENTION:  Thereof
;    NUMBER OF SEQUENCES:  37
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  No. 6117670artis Corporation
;      STREET:  3054 Cornwallis Road
;      CITY:  Research Triangle Park
;      STATE:  NC
;      COUNTRY:  USA
;      ZIP:  27709
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/028,934
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/729,214
;      FILING DATE:  09-OCT-1996
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/258,261
;      FILING DATE:  08-JUN-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Meigs, J. Timothy
;      REGISTRATION NUMBER:  38,241
;      REFERENCE/DOCKET NUMBER:  CGC1506/CIP7
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-541-8587
;      TELEFAX:  919-541-8689
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  28958 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
US-09-028-934-6

  Query Match             36.7%;  Score 348.2;  DB 3;  Length 28958;
  Best Local Similarity   62.7%;  Pred. No. 4.5e-57;
  Matches  564;  Conservative    0;  Mismatches  323;  Indels   12;  Gaps    1;

Qy          1 GTCTTCGTTTTTCCCGGGCAGGGCCCGCAATGGCCGGGCATGGGAAGGGAACTTCTCGAC 60

Db       9940 GTCTTCGTCTTTCCTGGGCAAGGCTCGCAGTGGGAAGGGATGGCCCTCTCCCTGCTCGAC 9999

Qy         61 GCTTCCGACGTCTTCCGGGAGAGCGTCCGCGCCTGCGAAGCCGCGTTCGCGCCCTACGTC 120

Db      10000 TCCTCGCCGGTCTTCCGCGCTCAGCTCGAAGCATGCGAGCGCGCGCTCGCTCCTCACGTC 10059

Qy        121 GACTGGTCGGTGGAGCAGGTGTTGCGGGACTCGCCGGACGCTCCCGGGCTGGACCGGGTG 180

Db      10060 GAGTGGAGCCTGCTCGCCGTCCTGCGCCGCGACGAGGGCGCCCCCTCCCTCGACCGCGTC 10119

Qy        181 GACGTCGTCCAGCCGACCCTGTTCGCCGTCATGATCTCCCTGGCCGCCCTCTGGCGCTCG 240

Db      10120 GACGTCGTACAGCCCGCCCTCTTTGCCGTCATGGTCTCCCTGGCCGCCCTCTGGCGCTCG 10179

Qy        241 CAAGGGGTCGAGCCGTGCGCGGTGCTGGGACACAGCCTGGGCGAGATCGCGGCAGCCCAC 300

Db      10180 CTCGGCGTCGAGCCCGCCGCCGTCGTCGGCCACAGCCAGGGCGAGATCGCCGCCGCCTTC 10239

Qy        301 GTCTCGGGAGGCCTGTCCCTGGCCGACGCCGCACGCGTGGTGACGCTTTGGAGCCAGGCA 360

Db      10240 GTCGCAGGCGCTCTCTCCCTCGAGGACGCGGCGCGCATCGCCGCCCTGCGCAGGAAAGCG 10299

Qy        361 CAGACCACCCTTGCCGGGACCGGCGCGCTCGTCTCCGTCGCCGCCACGCCGGATGAGCTC 420

Db      10300 CTCACCACCGTCGGCGGCAACGGCGGCATGGCCGCCGTCGAGCTCGGCGCCTCCGACCTC 10359

Qy        421 CTGCCCCGAATCGCTCCGTGGACCGAGGACAACCCGGCGCGGCTCGCCGTCGCAGCCGTC 480

Db      10360 CAGACCTACCTCGCTCCCTGGGGCGA-----------CAGGCTCTCCACCGCCGCCGTC 10407

Qy        481 AACGGACCCCGGAGCACAGTCGTTTCCGGTGCCCGCGAGGCCGTCGCGGACCTGGTGGCC 540

Db      10408 AACAGCCCCAGGGCTACCCTCGTATCCGGCGAGCCCGCCGCCGTCGACGCGCTGCTCGAC 10467

Qy        541 GACCTCACCGCCGCGCAGGTGCGCACGCGCATGATCCCGGTGGACGTTCCCGCCCACTCC 600

Db      10468 GTCCTCACCGCCACCAAGGTGTTCGCCCGCAAGATCCGCGTCGACTACGCCTCCCACTCC 10527

Qy        601 CCCCTGATGTACGCCATCGAGGAACGGGTCGTCAGCGGCCTGCTGCCCATCACCCCACGC 660

Db      10528 GCCCAGATGGACGCCGTCCAAGACGAGCTCGCCGCAGGTCTAGCCAACATCGCTCCTCGG 10587

Qy        661 CCCTCCCGCATCCCCTTCCACTCCTCGGTGACCGGCGGCCGCCTCGACACCCGCGAGCTA 720

Db      10588 ACGTGCGAGCTCCCTCTTTATTCGACCGTCACCGGCACCAGGCTCGACGGCTCCGAGCTC 10647

Qy        721 GACGCGGCGTACTGGTACCGCAACATGTCGAGCACGGTCCGGTTCGAGCCCGCCGCCCGG 780

Db      10648 GACGGCGCGTACTGGTATCGAAACCTCCGGCAAACCGTCCTGTTCTCGAGCGCGACCGAG 10707

Qy        781 CTGCTTCTGCAGCAGGGGCCCAAGACGTTCGTCGAGATGAGCCCGCACCCGGTGCTGACC 840

Db      10708 CGGCTCCTCGACGATGGGCATCGCTTCTCCGTCGAGGTCAGCCCCCATCCCGTGCTCACG 10767
```

```
Qy        841 ATGGGCCTCCAGGAGCTCGCCCCGGACCTGGGCGACACCACCGGCACCGCCGACACCGT 899
              | | |||||  |||  |  |  |  |       |    ||| |  || || | || |
Db      10768 CTCGCCCTCCGCGAGACCTGCGAGCGCTCACCGCTCGATCCCGTCGTCGTCGGCTCCAT 10826


RESULT 10
US-08-764-233A-1
; Sequence 1, Application US/08764233A
; Patent No. 5716849
;  GENERAL INFORMATION:
;    APPLICANT:  Ligon, James M.
;    APPLICANT:  Schupp, Thomas
;    APPLICANT:  Beck, James J.
;    APPLICANT:  Hill, Dwight S.
;    APPLICANT:  Neff, Snezanna
;    APPLICANT:  Ryals, John A.
;    TITLE OF INVENTION:  Genes For The Biosynthesis Of Soraphen
;    NUMBER OF SEQUENCES:  10
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Ciba-Geigy Corporation
;      STREET:  520 White Plains Road, P.O. Box 2005
;      CITY:  Tarrytown
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10591
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/764,233A
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/729,214
;      FILING DATE:  09-OCT-1996
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/258,261
;      FILING DATE:  08-JUN-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Meigs, J. Timothy
;      REGISTRATION NUMBER:  38,241
;      REFERENCE/DOCKET NUMBER:  1506/CIP6
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (919) 541-8587
;      TELEFAX:  (919) 541-8689
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  49377 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    ORIGINAL SOURCE:
;      ORGANISM:  Sorangium cellulosum
;    IMMEDIATE SOURCE:
;      CLONE:  p98/1, pJL3, and pVKM15
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  383..760
;      OTHER INFORMATION:  /product= "SorR"
;      OTHER INFORMATION:  /note= "This gene encodes a protein that is highly homologous
;      OTHER INFORMATION:  the reductase domains of type I PKSs such as eryA from
;      OTHER INFORMATION:  Saccharopolyspora erythraea."
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  927..19874
;      OTHER INFORMATION:  /product= "SorA"
;      OTHER INFORMATION:  /note= "Gene product is highly homologous to type I PKSs that
;      OTHER INFORMATION:  are known to be involved in the synthesis of polyketide
;      OTHER INFORMATION:  compounds."
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  942..7115
;      OTHER INFORMATION:  /product= "Module 1 of SorA"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  7203..12884
;      OTHER INFORMATION:  /product= "Module 2 of SorA"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  13455..19616
;      OTHER INFORMATION:  /product= "Module 3 of SorA"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  19871..46318
;      OTHER INFORMATION:  /product= "SorB"
;      OTHER INFORMATION:  /note= "Gene product is highly homologous to type I PKS gene
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  19870..24556
;      OTHER INFORMATION:  /product= "Module 1 of SorB"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  24638..30820
;      OTHER INFORMATION:  /product= "Module 2 of SorB"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  30881..35446
;      OTHER INFORMATION:  /product= "Module 3 of SorB"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  35528..40114
;      OTHER INFORMATION:  /product= "Module 4 of SorB"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  40190..46318
;      OTHER INFORMATION:  /product= "Module 5 of SorB"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  46851..47891
;      OTHER INFORMATION:  /product= "SorM"
;      OTHER INFORMATION:  /note= "The protein encoded by the sorM gene is highly
;      OTHER INFORMATION:  homologous to the methyltransferase from Streptomyces
;      OTHER INFORMATION:  hygroscopicus that is involved in the synthesis of the
;      OTHER INFORMATION:  polyketide rappamicin."
US-08-764-233A-1

  Query Match             36.7%;  Score 348.2;  DB 1;  Length 49377;
  Best Local Similarity   62.7%;  Pred. No. 4.6e-57;
  Matches  564;  Conservative    0;  Mismatches  323;  Indels   12;  Gaps    1;

Qy          1 GTCTTCGTTTTTCCCGGGCAGGGCCCGCAATGGCCGGGCATGGGAAGGGAACTTCTCGAC 60
              |||||||| ||||| ||||| ||| |||| |||   || ||||          || ||||||
Db      26243 GTCTTCGTCTTTCCTGGGCAAGGCTCGCAGTGGGAAGGGATGGCCCTCTCCCTGCTCGAC 26302

Qy         61 GCTTCCGACGTCTTCCGGGAGAGCGTCCGCGCCTGCGAAGCCGCGTTCGCGCCCTACGTC 120
              | ||    |||||||| |       ||   || |||||   |||| |||| ||  |||||
Db      26303 TCCTCGCCGGTCTTCCGCGCTCAGCTCGAAGCATGCGAGCGCGCGCTCGCTCCTCACGTC 26362

Qy        121 GACTGGTCGGTGGAGCAGGTGTTGCGGGACTCGCCGGACGCTCCCGGGCTGGACCGGGTG 180
              || |||    ||       ||  ||||   |      || ||| |||   || ||||| ||
Db      26363 GAGTGGAGCCTGCTCGCCGTCCTGCGCCGCGACGAGGGCGCCCCCTCCCTCGACCGCGTC 26422

Qy        181 GACGTCGTCCAGCCGACCCTGTTCGCCGTCATGATCTCCCTGGCCGCCCTCTGGCGCTCG 240
              |||||||| |||||  |||| || |||||||| ||||||||||||||||||||||||||
Db      26423 GACGTCGTACAGCCCGCCCTCTTTGCCGTCATGGTCTCCCTGGCCGCCCTCTGGCGCTCG 26482

Qy        241 CAAGGGGTCGAGCCGTGCGCGGTGCTGGGACACAGCCTGGGCGAGATCGCGGCAGCCCAC 300
              |  || ||||||||   ||| ||  | || ||||||| |||||||||||| || |||  |
Db      26483 CTCGGCGTCGAGCCCGCCGCCGTCGTCGGCCACAGCCAGGGCGAGATCGCCGCCGCCTTC 26542

Qy        301 GTCTCGGGAGGCCTGTCCCTGGCCGACGCCGCACGCGTGGTGACGCTTTGGAGCCAGGCA 360
              ||| | || |  || ||||| |  ||||| || ||| | |   | ||  | ||  | ||
Db      26543 GTCGCAGGCGCTCTCTCCCTCGAGGACGCGGCGCGCATCGCCGCCCTGCGCAGGAAAGCG 26602
```

```
Qy        361 CAGACCACCCTTGCCGGGACCGGCGCGCTCGTCTCCGTCGCCGCCACGCCGGATGAGCTC 420
Db      26603 CTCACCACCGTCGGCGGCAACGGCGGCATGGCCGCCGTCGAGCTCGGCGCCTCCGACCTC 26662

Qy        421 CTGCCCCGAATCGCTCCGTGGACCGAGGACAACCCGGCGCGGCTCGCCGTCGCAGCCGTC 480
Db      26663 CAGACCTACCTCGCTCCCTGGGGCGA----------CAGGCTCTCCACCGCCGCCGTC 26710

Qy        481 AACGGACCCCGGAGCACAGTCGTTTCCGGTGCCCGCGAGGCCGTCGCGGACCTGGTGGCC 540
Db      26711 AACAGCCCCAGGGCTACCCTCGTATCCGGCGAGCCCGCCGCCGTCGACGCGCTGCTCGAC 26770

Qy        541 GACCTCACCGCCGCGCAGGTGCGCACGCGCATGATCCCGGTGGACGTTCCCGCCCACTCC 600
Db      26771 GTCCTCACCGCCACCAAGGTGTTCGCCCGCAAGATCCGCGTCGACTACGCCTCCCACTCC 26830

Qy        601 CCCCTGATGTACGCCATCGAGGAACGGGTCGTCAGCGGCCTGCTGCCCATCACCCCACGC 660
Db      26831 GCCCAGATGGACGCCGTCCAAGACGAGCTCGCCGCAGGTCTAGCCAACATCGCTCCTCGG 26890

Qy        661 CCCTCCCGCATCCCCTTCCACTCCTCGGTGACCGGCGGCCGCCTCGACACCCGCGAGCTA 720
Db      26891 ACGTGCGAGCTCCCTCTTTATTCGACCGTCACCGGCACCAGGCTCGACGGCTCCGAGCTC 26950

Qy        721 GACGCGGCGTACTGGTACCGCAACATGTCGAGCACGGTCCGGTTCGAGCCCGCCGCCCGG 780
Db      26951 GACGGCGCGTACTGGTATCGAAACCTCCGGCAAACCGTCCTGTTCTCGAGCGCGACCGAG 27010

Qy        781 CTGCTTCTGCAGCAGGGGCCCAAGACGTTCGTCGAGATGAGCCCGCACCCGGTGCTGACC 840
Db      27011 CGGCTCCTCGACGATGGGCATCGCTTCTCCGTCGAGGTCAGCCCCCATCCCGTGCTCACG 27070

Qy        841 ATGGGCCTCCAGGAGCTCGCCCCGGACCTGGGCGACACCACCGGCACCGCCGACACCGT 899
Db      27071 CTCGCCCTCCGCGAGACCTGCGAGCGCTCACCGCTCGATCCCGTCGTCGTCGGCTCCAT 27129
```

```
RESULT 11
US-08-804-227C-7
; Sequence 7, Application US/08804227C
; Patent No. 5876991
;  GENERAL INFORMATION:
;    APPLICANT:  DeHoff, Bradley S.
;    APPLICANT:  Kuhstoss, Stuart A.
;    APPLICANT:  Rosteck, Paul R., Jr.
;    APPLICANT:  Sutton, Kimberly L.
;    TITLE OF INVENTION:  POLYKETIDE SYNTHASE GENES
;    NUMBER OF SEQUENCES:  15
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  THOMAS G. PLANT  1501
;      STREET:  LILLY CORPORATE CENTER
;      CITY:  INDIANAPOLIS
;      STATE:  IN
;      COUNTRY:  USA
;      ZIP:  46285
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  MS-DOS
;      SOFTWARE:  ASCI(DOS) Text only
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/804,227C
;      FILING DATE:  February 21, 1997
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Plant, Thomas, G.
;      REGISTRATION NUMBER:  35,784
;      REFERENCE/DOCKET NUMBER:  X-8231
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  317-276-2459
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  44377 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  350..14002
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  14046..20036
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  20110..31284
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  31329..36071
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  36155..41830
US-08-804-227C-7

  Query Match             36.3%;  Score 344;  DB 2;  Length 44377;
  Best Local Similarity   62.5%;  Pred. No. 2.8e-56;
  Matches  599;  Conservative    0;  Mismatches  335;  Indels   24;  Gaps    3;

Qy          1 GTCTTCGTTTTTCCCGGGCAGGGCCCGCAATGGCCGGGCATGGGAAGGGAACTTCT---- 56
Db      26374 GTGTTCGTCTTCCCCGGCCAGGGCTCGCAGTGGCCCGAGATGGCCGACGGGCTGCTGGCC 26433

Qy         57 -----CGACGCTTCCGACGTCTTCCGGGAGAGCGTCCGCGCCTGCGAAGCCGCGTTCGCG 111
Db      26434 CGCTCCAGCGGCTCCGGCTCCTTCCTGGAGACCGCCCGCGCCTGCGACCTCGCGCTCCGG 26493

Qy        112 CCCTACGTCGACTGGTCGGTGGAGCAGGTGTTGCGGGACTCGCCGGACGCTCCCGGGCTG 171
Db      26494 CCCCACCTCGGCTGGTCCGTCCTGGACGTACTGCGCCGGGAACCCGGCGCGCCCTCGCTC 26553

Qy        172 GACCGGGTGGACGTCGTCCAGCCGACCCTGTTCGCCGTCATGATCTCCCTGGCCGCCCTC 231
Db      26554 GACCGGGTCGACGTGGTGCAGCCCGTGCTGTTCACCATGATGGTCTCGCTCGCCGAGACG 26613

Qy        232 TGGCGCTCGCAAGGGGTCGAGCCGTGCGCGGTGCTGGGACACAGCCTGGGCGAGATCGCG 291
Db      26614 TGGCGTTCGCTGGGCGTCGAACCGGCCGCGGTCGTCGGTCACTCCCAGGGCGAGATCGCC 26673

Qy        292 GCAGCCCACGTCTCGGGAGGCCTGTCCCTGGCCGACGCCGCACGCGTGGTGACGCTTTGG 351
Db      26674 GCCGCCTACGTCGCCGGCGCCCTGACGCTGGACGACGCGGCGCGCATCGTCGCCCTGCGC 26733

Qy        352 AGCCAGGCACAGACCACCCTTGCCGGGACCGGCGCGCTCGTCTCCGTCGCCGCCACGCCG 411
Db      26734 AGCCAGGCGTGGCTGCGGCTGGCCGGCAAGGGCGGCATGGTCGCCGTGACCCTGTCCGAA 26793

Qy        412 GATGAGCTCCTGCCCCGAATCGCTCCGTGGACCGAGGACAACCCGGCGCGGCTCGCCGTC 471
Db      26794 CGCGACCTGCGTCCCCGCCTGGAGCCCTGGAGCGA----------CCGGCTCGCCGTC 26841

Qy        472 GCAGCCGTCAACGGACCCCGGAGCACAGTCGTTTCCGGTGCCCGCGAGGCCGTCGCGGAC 531
Db      26842 GCCGCCGTCAACGGCCCCGAGACCTGCGCCGTCTCCGGGGACCCGGACGCCCTGGCGGAG 26901

Qy        532 CTGGTGGCCGACCTCACCGCCGCGCAGGTGCGCACGCGCATGATCCCGGTGGACGTTCCC 591
Db      26902 CTGGTCGCCGAACTCGGTGCGGAGGGCGTGCACGCCCGCCCCATCCCCGGCGTCGACACC 26961

Qy        592 GCC---CACTCCCCCCTGATGTACGCCATCGAGGAACGGGTCGTCAGCGGCCTGCTGCCC 648
Db      26962 GCCGGGCACTCGCCGCAGGTCGACACGCTGGAGGCCCACCTGCGGAAGGTGCTCGCGCCC 27021

Qy        649 ATCACCCCACGCCCCTCCCGCATCCCCTTCCACTCCTCGGTGACCGGCGGCCGCCTCGAC 708
Db      27022 GTCGCGCCCCGCACCTCCGACATCCCGTTCTACTCGACGGTCACCGGAGGACTGATCGAC 27081
```

```
Qy        709 ACCCGCGAGCTAGACGCGGCGTACTGGTACCGCAACATGTCGAGCACGGTCCGGTTCGAG 768
              |||  |||||| |||||  |  ||||||||||||||||||        ||||   |||||||
Db      27082 ACCGCCGAGCTGGACGCCGACTACTGGTACCGCAACATGCGCGAGCCGGTGGAGTTCGAG 27141

Qy        769 CCCGCCGCCCGGCTGCTTCTGCAGCAGGGGCCCAAGACGTTCGTCGAGATGAGCCCGCAC 828
              |  ||| ||||    ||  |      | || | | |    |||| | |||  ||||||||||
Db      27142 CAGGCCACCCGCGCCCTGATCGCCGACGGCCACGACGTGTTCCTGGAGTCGAGCCCGCAC 27201

Qy        829 CCGGTGCTGACCATGGGCCTCCAGGAGCTCGCCCCGGACCTGGGCGACACCACCGGCACC 888
              ||  ||||| || |   |||||||||      |  |||  ||  | | | | | |  |
Db      27202 CCCATGCTGGCCGTCTCCCTCCAGGAGACGATCAGCGACGCCGGTTCCCCGGCGGCCGTC 27261

Qy        889 GCCGACACCGTGATCATGGGCACGCTGCGCCGCGGCCAGGGCACCCTGGACCACTTCC 946
                || |||| ||      |||  |    ||| | ||  |      | | | || || ||
Db      27262 CTCGGCACCCTGCGGCGCGGCCAGGGCGGCCCCCGCTGGCTGGGCGTCGCCCTCTGCC 27319


RESULT 12
US-08-804-198-1
; Sequence 1, Application US/08804198
; Patent No. 5945320
;  GENERAL INFORMATION:
;    APPLICANT:  Burgett, Stanley G.
;    APPLICANT:  Kuhstoss, Stuart A.
;    APPLICANT:  Rao, Nagaraja R.
;    APPLICANT:  Richardson, Mark A.
;    APPLICANT:  Rosteck, Paul R., Jr.
;    TITLE OF INVENTION:  PLATENOLIDE SYNTHASE GENE
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  PAUL R. CANTRELL  1138
;      STREET:  LILLY CORPORATE CENTER
;      CITY:  INDIANAPOLIS
;      STATE:  IN
;      COUNTRY:  USA
;      ZIP:  46285
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  Macintosh
;      OPERATING SYSTEM:  Macintosh 7.0
;      SOFTWARE:  Microsoft Word 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/804,198
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  CANTRELL, PAUL R.
;      REGISTRATION NUMBER:  36,470
;      REFERENCE/DOCKET NUMBER:  P9113
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  317-276-3885
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  44377 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic).
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  350..14002
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  14046..20036
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  20110..31284
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  31329..36071
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  36155..41830
US-08-804-198-1

  Query Match             36.3%;  Score 344;  DB 2;  Length 44377;
  Best Local Similarity   62.5%;  Pred. No. 2.8e-56;
  Matches  599;  Conservative    0;  Mismatches  335;  Indels   24;  Gaps    3;

Qy          1 GTCTTCGTTTTTCCCGGGCAGGGCCCGCAATGGCCGGGCATGGGAAGGGAACTTCT---- 56
              || ||||| || ||||| |||||| |||| ||||| |  ||||      |  || ||
Db      26374 GTGTTCGTCTTCCCCGGCCAGGGCTCGCAGTGGCCCGAGATGGCCGACGGGCTGCTGGCC 26433

Qy         57 -----CGACGCTTCCGACGTCTTCCGGGAGAGCGTCCGCGCCTGCGAAGCCGCGTTCGCG 111
                  |  ||  |||| |  ||||| ||||| || ||||||||||||    |||| ||  |
Db      26434 CGCTCCAGCGGCTCCGGCTCCTTCCTGGAGACCGCCCGCGCCTGCGACCTCGCGCTCCGG 26493

Qy        112 CCCTACGTCGACTGGTCGGTGGAGCAGGTGTTGCGGGACTCGCCGGACGCTCCCGGGCTG 171
              ||| || ||| |||||| ||  | | ||  ||||      || | ||| |||  |||
Db      26494 CCCCACCTCGGCTGGTCCGTCCTGGACGTACTGCGCCGGGAACCCGGCGCGCCCTCGCTC 26553

Qy        172 GACCGGGTGGACGTCGTCCAGCCGACCCTGTTCGCCGTCATGATCTCCCTGGCCGCCCTC 231
              |||||||| ||||| || |||||     ||||||  || | |||  |||| || ||||
Db      26554 GACCGGGTCGACGTGGTGCAGCCCGTGCTGTTCACCATGATGGTCTCGCTCGCCGAGACG 26613

Qy        232 TGGCGCTCGCAAGGGGTCGAGCCGTGCGCGGTGCTGGGACACAGCCTGGGCGAGATCGCG 291
              |||||  ||||  || |||||  |||  ||||||  | ||  |||   || ||||||||||
Db      26614 TGGCGTTCGCTGGGCGTCGAACCGGCCGCGGTCGTCGGTCACTCCCAGGGCGAGATCGCC 26673

Qy        292 GCAGCCCACGTCTCGGGAGGCCTGTCCCTGGCCGACGCCGCACGCGTGGTGACGCTTTGG 351
              || ||| ||||| | || | |||| | |||| |||||| || ||| | ||  | ||  |
Db      26674 GCCGCCTACGTCGCCGGCGCCCTGACGCTGGACGACGCGGCGCGCATCGTCGCCCTGCGC 26733

Qy        352 AGCCAGGCACAGACCACCCTTGCCGGGACCGGCGCGCTCGTCTCCGTCGCCGCCACGCCG 411
              ||||||||    |       || ||||| |  ||||    | ||| ||||   ||     |
Db      26734 AGCCAGGCGTGGCTGCGGCTGGCCGGCAAGGGCGGCATGGTCGCCGTGACCCTGTCCGAA 26793

Qy        412 GATGAGCTCCTGCCCCGAATCGCTCCGTGGACCGAGGACAACCCGGCGCGGCTCGCCGTC 471
                || || |  ||||| . | |  || |||| |||                ||||||||||
Db      26794 CGCGACCTGCGTCCCCGCCTGGAGCCCTGGAGCGA-----------CCGGCTCGCCGTC 26841

Qy        472 GCAGCCGTCAACGGACCCCGGAGCACAGTCGTTTCCGGTGCCCGCGAGGCCGTCGCGGAC 531
              || |||||||||| ||| || |   | ||| |||| | ||  || ||| | |||||
Db      26842 GCCGCCGTCAACGGCCCCGAGACCTGCGCCGTCTCCGGGGACCCGGACGCCCTGGCGGAG 26901

Qy        532 CTGGTGGCCGACCTCACCGCCGCGCAGGTGCGCACGCGCATGATCCCGGTGGACGTTCCC 591
              ||||| |||| ||| || | |  |||| | | |||   |||||  | | ||   ||
Db      26902 CTGGTCGCCGAACTCGGTGCGGAGGGCGTGCACGCCCGCCCCATCCCCGGCGTCGACACC 26961

Qy        592 GCC---CACTCCCCCCTGATGTACGCCATCGAGGAACGGGTCGTCAGCGGCCTGCTGCCC 648
              |||   ||||| || | | |  || | | |||  |   |   | | ||  ||||
Db      26962 GCCGGGCACTCGCCGCAGGTCGACACGCTGGAGGCCCACCTGCGGAAGGTGCTCGCGCCC 27021

Qy        649 ATCACCCCACGCCCCTCCCGCATCCCCTTCCACTCCTCGGTGACCGGCGGCCGCCTCGAC 708
              || | || ||| ||||| |||||| ||| |||| |||| ||||| || |   |||||
Db      27022 GTCGCGCCCCGCACCTCCGACATCCCGTTCTACTCGACGGTCACCGGAGGACTGATCGAC 27081

Qy        709 ACCCGCGAGCTAGACGCGGCGTACTGGTACCGCAACATGTCGAGCACGGTCCGGTTCGAG 768
              |||  |||||| |||||  |  ||||||||||||||||||        ||||   |||||||
Db      27082 ACCGCCGAGCTGGACGCCGACTACTGGTACCGCAACATGCGCGAGCCGGTGGAGTTCGAG 27141

Qy        769 CCCGCCGCCCGGCTGCTTCTGCAGCAGGGGCCCAAGACGTTCGTCGAGATGAGCCCGCAC 828
              |  ||| ||||    ||  |      | || | | |    |||| | |||  ||||||||||
Db      27142 CAGGCCACCCGCGCCCTGATCGCCGACGGCCACGACGTGTTCCTGGAGTCGAGCCCGCAC 27201

Qy        829 CCGGTGCTGACCATGGGCCTCCAGGAGCTCGCCCCGGACCTGGGCGACACCACCGGCACC 888
              ||  ||||| || |   |||||||||      |  |||  ||  | | | | | |  |
Db      27202 CCCATGCTGGCCGTCTCCCTCCAGGAGACGATCAGCGACGCCGGTTCCCCGGCGGCCGTC 27261

Qy        889 GCCGACACCGTGATCATGGGCACGCTGCGCCGCGGCCAGGGCACCCTGGACCACTTCC 946
                || |||| ||      |||  |    ||| | ||  |      | | | || || ||
Db      27262 CTCGGCACCCTGCGGCGCGGCCAGGGCGGCCCCCGCTGGCTGGGCGTCGCCCTCTGCC 27319


RESULT 13
```

```
US-08-858-003-29
; Sequence 29, Application US/08858003
; Patent No. 6060234
;  GENERAL INFORMATION:
;    APPLICANT:  Katz, Leonard
;    APPLICANT:  Stassi, Diane L.
;    APPLICANT:  Summers Jr., Richard G.
;    APPLICANT:  Ruan, Xiaoan
;    APPLICANT:  Pereda-Lopez, Ana
;    APPLICANT:  Kakavas, Stephan J.
;    TITLE OF INVENTION:  NOVEL POLYKETIDE DERIVATIVES
;    TITLE OF INVENTION:  AND RECOMBINANT METHODS FOR MAKING SAME
;    NUMBER OF SEQUENCES:  34
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Abbott Laboratories
;      STREET:  100 Abbott Park Rd.
;      CITY:  Abbott Park
;      STATE:  Illinois
;      COUNTRY:  USA
;      ZIP:  60064-3500
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/858,003
;      FILING DATE:  16-MAY-1979
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Dianne Casuto
;      REGISTRATION NUMBER:  P-40,943
;      REFERENCE/DOCKET NUMBER:  4952.US.P2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (847)-938-3137
;      TELEFAX:  (847)-938-2623
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  29:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1010 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
US-08-858-003-29

  Query Match           35.6%;  Score 337.4;  DB 3;  Length 1010;
  Best Local Similarity   61.4%;  Pred. No. 4.5e-55;
  Matches  583;  Conservative    0;  Mismatches  351;  Indels   15;  Gaps    2;

Qy          1 GTCTTCGTTTTTCCCGGGCAGGGCCCGCAATGGCCGGGCATGGGAAGGGAACTTCTCGAC 60
              || ||||| || ||||| |||||| |||| ||| | || ||||    ||  || || ||
Db         13 GTGTTCGTGTTCCCCGGCCAGGGCTCGCAGTGGGCCGGAATGGCCGAGGGGCTGCTGGAG 72

Qy         61 GCTTCCGACGTCTTCCGGGAGAGCGTCCGCGCCTGCGAAGCCGCGTTCGCGCCCTACGTC 120
                |||| ||  ||||||      | |  | | ||||| |||||| |   ||| ||| ||
Db         73 CGGTCCGGCGCGTTCCGGAGTGCGGCCGACTCGTGCGACGCCGCGCTGCGGCCGTACCTC 132

Qy        121 GACTGGTCGGTGGAGCAGGTGTTGCGGGACTCGCCGGACGCTCCCGGGCTGGACCGGGTG 180
              | ||||||||||  |   ||| |||| |      |||||||| |||  ||| ||||||||
Db        133 GGCTGGTCGGTGCTGAGCGTGCTGCGCGGGGAACCGGACGCGCCCTCGCTCGACCGGGTC 192

Qy        181 GACGTCGTCCAGCCGACCCTGTTCGCCGTCATGATCTCCCTGGCCGCCCTCTGGCGCTCG 240
              |||||||| ||||||   |||||| |  | ||| |||| || || ||  |||||||  ||
Db        193 GACGTCGTGCAGCCGGTGCTGTTCACGATGATGGTCTCGCTCGCGGCGGTCTGGCGTGCG 252

Qy        241 CAAGGGGTCGAGCCGTGCGCGGTGCTGGGACACAGCCTGGGCGAGATCGCGGCAGCCCAC 300
              |  ||||| || |||   |||||  | || |||   | ||| ||||||||| || |||||
Db        253 CTGGGGGTGGAACCGGCGGCGGTCGTCGGGCACTCGCAGGGTGAGATCGCCGCTGCCCAT 312

Qy        301 GTCTCGGGAGGCCTGTCCCTGGCCGACGCCGCACGCGTGGTGACGCTTTGGAGCCAGGCA 360
              ||| | || |  |||||| |||| |||| | || ||  | ||  | ||  | || | |||
Db        313 GTCGCCGGTGCGCTGTCGCTGGACGACTCGGCCCGGATCGTCGCCCTGCGCAGTCGGGCG 372

Qy        361 CAGACCACCCTTGCCGGGACCGGCGCGCTCGTCTCCGTCGCCGCCACGCCGGATGAGCTC 420
              |  |   || || || |  ||||   | ||  | ||  |      || |||| |||||
Db        373 TGGCTCGGACTGGCGGGCAAGGGCGGCATGGTGGCGGTGCCGATGCCGGCGGAGGAGCTG 432

Qy        421 CTGCCCCGAATCGCTCCGTGGACCGAGGACAACCCGGCGCGGCTCGCCGTCGCAGCCGTC 480
              | ||| ||  | |   |||||                ||  || || |||||||| ||||||
Db        433 CGGCCGCGGCTGGTGACGTGG------------GGGGACCGTCTGGCCGTCGCCGCCGTC 480

Qy        481 AACGGACCCCGGAGCACAGTCGTTTCCGGTGCCCGCGAGGCCGTCGCGGACCTGGTGGCC 540
              ||| | ||| |   |   | |||  | || | ||  ||||  | || || ||||||||
Db        481 AACAGCCCCGGTTCCTGCGCCGTCGCAGGCGACCCGGAGGCGCTGGCCGAACTGGTGGCG 540

Qy        541 GACCTCACCGCCGCGCAGGTGCGCACGCGCATGATCCCGGTGGACGTTCCCGC---CCAC 597
              || |||| | | |||||  | | ||   |||||| |  | ||   | ||    ||||
Db        541 CTGCTGACCGGTGAGGGGGTGCACGCCCGGCCGATCCCCGGCGTCGACACGGCGGGCCAC 600

Qy        598 TCCCCCCTGATGTACGCCATCGAGGAACGGGTCGTCAGCGGCCTGCTGCCCATCACCCCA 657
              || || | | || |||  |   ||  |   |  |     |  |||   ||  || ||||
Db        601 TCGCCGCAGGTGGACGCGTTGCGGGCTCATCTGCTGGAGGTGCTGGCCCCGGTCGCCCCC 660

Qy        658 CGCCCCTCCCGCATCCCCTTCCACTCCTCGGTGACCGGCGGCCGCCTCGACACCCGCGAG 717
              || ||  ||  |||||| ||| ||||  ||||||||||||| |  ||  |||  |  ||||
Db        661 CGACCGGCCGACATCCCGTTCTACTCGACGGTGACCGGCGGGCTGCTGGACGGCACCGAG 720

Qy        718 CTAGACGCGGCGTACTGGTACCGCAACATGTCGAGCACGGTCCGGTTCGAGCCCGCCGCC 777
              || |||||| |||||||||||||||||||        | |||  ||||||||  ||| |
Db        721 CTGGACGCGACGTACTGGTACCGCAACATGCGCGAGCCCGTCGAGTTCGAGCGGGCCACA 780

Qy        778 CGGCTGCTTCTGCAGCAGGGGCCCAAGACGTTCGTCGAGATGAGCCCGCACCCGGTGCTG 837
              |||  |||  |      | |||| | |    |:| | |||| |||||||| ||  |||||
Db        781 CGGGCGCTGATCGCCGACGGGCACGACGTCTTCCTGGAGACGAGCCCGCATCCCATGCTG 840

Qy        838 ACCATGGGCCTCCAGGAGCTCGCCCCGGACCTGGGCGACACCACCGGCACCGCCGACACC 897
              || |||  ||  || ||   | | | |||   |||  |  | | |       ||  |||
Db        841 GCCGTGGCGCTGGAGCAGACGGTCACCGACGCCGGCACCGACGCGGCGGTGCTCGGGACC 900

Qy        898 GTGATCATGGGCACGCTGCGCCGCGGCCAGGGCACCCTGGACCACTTCC 946
              ||  |   ||        | |   ||  |   ||||| |  || ||
Db        901 CTGCGCCGCCGCCACGGCGGTCCTCGCGCGCTGGCCCTGGCCGTCTGCC 949


RESULT 14
US-09-078-166-29
; Sequence 29, Application US/09078166
; Patent No. 6063561
;  GENERAL INFORMATION:
;    APPLICANT:  Katz, Leonard
;    APPLICANT:  Stassi, Diane L.
;    APPLICANT:  Summers Jr., Richard G.
;    APPLICANT:  Ruan, Xiaoan
;    APPLICANT:  Pereda-Lopez, Ana
;    APPLICANT:  Kakavas, Stephan J.
;    TITLE OF INVENTION:  NOVEL POLYKETIDE DERIVATIVES
;    TITLE OF INVENTION:  AND RECOMBINANT METHODS FOR MAKING SAME
;    NUMBER OF SEQUENCES:  44
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Abbott Laboratories
;      STREET:  100 Abbott Park Rd.
;      CITY:  Abbott Park
;      STATE:  Illinois
;      COUNTRY:  USA
;      ZIP:  60064-3500
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/078,166
;      FILING DATE:  16-MAY-1979
```

```
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Dianne Casuto
;      REGISTRATION NUMBER:  P-40,943
;      REFERENCE/DOCKET NUMBER:  4952.US.P2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (847)-938-3137
;      TELEFAX:  (847)-938-2623
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  29:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1010 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
US-09-078-166-29

  Query Match           35.6%;  Score 337.4;  DB 3;  Length 1010;
  Best Local Similarity   61.4%;  Pred. No. 4.5e-55;
  Matches  583;  Conservative    0;  Mismatches  351;  Indels   15;  Gaps    2;

Qy          1 GTCTTCGTTTTTCCCGGGCAGGGCCCGCAATGGCCGGGCATGGGAAGGGAACTTCTCGAC 60
              || ||||| || ||||| |||||| |||| ||| | || ||||    ||  || || ||
Db         13 GTGTTCGTGTTCCCCGGCCAGGGCTCGCAGTGGGCCGGAATGGCCGAGGGGCTGCTGGAG 72

Qy         61 GCTTCCGACGTCTTCCGGGAGAGCGTCCGCGCCTGCGAAGCCGCGTTCGCGCCCTACGTC 120
                |||| ||  ||||||       | |  | | ||||| |||||| |   ||| ||| ||
Db         73 CGGTCCGGCGCGTTCCGGAGTGCGGCCGACTCGTGCGACGCCGCGCTGCGGCCGTACCTC 132

Qy        121 GACTGGTCGGTGGAGCAGGTGTTGCGGGACTCGCCGGACGCTCCCGGGCTGGACCGGGTG 180
              | ||||||||||  |   ||| |||| |      |||||||| |||  ||| ||||||||
Db        133 GGCTGGTCGGTGCTGAGCGTGCTGCGCGGGGAACCGGACGCGCCCTCGCTCGACCGGGTC 192

Qy        181 GACGTCGTCCAGCCGACCCTGTTCGCCGTCATGATCTCCCTGGCCGCCCTCTGGCGCTCG 240
              |||||||| ||||||    |||||| |  | ||| |||| || || ||  |||||||  ||
Db        193 GACGTCGTGCAGCCGGTGCTGTTCACGATGATGGTCTCGCTCGCGGCGGTCTGGCGTGCG 252

Qy        241 CAAGGGGTCGAGCCGTGCGCGGTGCTGGGACACAGCCTGGGCGAGATCGCGGCAGCCCAC 300
              |  ||||| || |||    |||||  | || |||    | ||| |||||||| || |||||
Db        253 CTGGGGGTGGAACCGGCGGCGGTCGTCGGGCACTCGCAGGGTGAGATCGCCGCTGCCCAT 312

Qy        301 GTCTCGGGAGGCCTGTCCCTGGCCGACGCCGCACGCGTGGTGACGCTTTGGAGCCAGGCA 360
              ||| | || |  ||||| |||| |||| | || ||  | ||  | ||  | || | |||
Db        313 GTCGCCGGTGCGCTGTCGCTGGACGACTCGGCCCGGATCGTCGCCCTGCGCAGTCGGGCG 372

Qy        361 CAGACCACCCTTGCCGGGACCGGCGCGCTCGTCTCCGTCGCCGCCACGCCGGATGAGCTC 420
                |  |   || || || |  ||||   | ||  | ||  |      || |||| |||||
Db        373 TGGCTCGGACTGGCGGGCAAGGGCGGCATGGTGGCGGTGCCGATGCCGGCGGAGGAGCTG 432

Qy        421 CTGCCCCGAATCGCTCCGTGGACCGAGGACAACCCGGCGCGGCTCGCCGTCGCAGCCGTC 480
              | ||| ||  | |  |||||             ||  || || |||||||| ||||||
Db        433 CGGCCGCGGCTGGTGACGTGG------------GGGGACCGTCTGGCCGTCGCCGCCGTC 480

Qy        481 AACGGACCCCGGAGCACAGTCGTTTCCGGTGCCCGCGAGGCCGTCGCGGACCTGGTGGCC 540
              ||| | |||  |   |  | |||  | || | ||  |||||  | ||  || ||||||||
Db        481 AACAGCCCCGGTTCCTGCGCCGTCGCAGGCGACCCGGAGGCGCTGGCCGAACTGGTGGCG 540

Qy        541 GACCTCACCGCCGCGCAGGTGCGCACGCGCATGATCCCGGTGGACGTTCCCGC---CCAC 597
                || ||||  | |  ||||| | | ||   ||||||  |  | ||   | ||   ||||
Db        541 CTGCTGACCGGTGAGGGGGTGCACGCCCGGCCGATCCCCGGCGTCGACACGGCGGGCCAC 600

Qy        598 TCCCCCCTGATGTACGCCATCGAGGAACGGGTCGTCAGCGGCCTGCTGCCCCATCACCCCA 657
              || || | | || ||||  |   ||  |   | |    |  |||   ||  || ||||
Db        601 TCGCCGCAGGTGGACGCGTTGCGGGCTCATCTGCTGGAGGTGCTGGCCCCGGTCGCCCCC 660

Qy        658 CGCCCCTCCCGCATCCCCTTCCACTCCTCGGTGACCGGCGGCCGCCTCGACACCCGCGAG 717
              || ||  ||  |||||| ||| ||||  |||||||||||| |  || |||  |  ||||
Db        661 CGACCGGCCGACATCCCGTTCTACTCGACGGTGACCGGCGGGCTGCTGGACGGCACCGAG 720

Qy        718 CTAGACGCGGCGTACTGGTACCGCAACATGTCGAGCACGGTCCGGTTCGAGCCCGCCGCC 777
              || |||||| |||||||||||||||||||        | |||  ||||||||  ||| |
Db        721 CTGGACGCGACGTACTGGTACCGCAACATGCGCGAGCCCGTCGAGTTCGAGCGGGCCACA 780

Qy        778 CGGCTGCTTCTGCAGCAGGGGCCCAAGACGTTCGTCGAGATGAGCCCGCACCCGGTGCTG 837
              |||  |||  |      | |||| | |    ||| | |||| ||||||||| ||  |||||
Db        781 CGGGCGCTGATCGCCGACGGGCACGACGTCTTCCTGGAGACGAGCCCGCATCCCATGCTG 840

Qy        838 ACCATGGGCCTCCAGGAGCTCGCCCCGGACCTGGGCGACACCACCGGCACCGCCGACACC 897
              || |||  ||  || ||   | | | |||  ||| |  | | |        || |||.
Db        841 GCCGTGGCGCTGGAGCAGACGGTCACCGACGCCGGCACCGACGCGGCGGTGCTCGGGACC 900

Qy        898 GTGATCATGGGCACGCTGCGCCGCGGCCAGGGCACCCTGGACCACTTCC 946
              ||  |    ||         | |    ||  |    ||||||  |  || ||
Db        901 CTGCGCCGCCGCCACGGCGGTCCTCGCGCGCTGGCCCTGGCCGTCTGCC 949


RESULT 15
US-08-997-467-29
; Sequence 29, Application US/08997467
; Patent No. 6200813
;  GENERAL INFORMATION:
;    APPLICANT:  Katz, Leonard
;    APPLICANT:  Stassi, Diane L.
;    APPLICANT:  Summers Jr., Richard G.
;    APPLICANT:  Ruan, Xiaoan
;    APPLICANT:  Pereda-Lopez, Ana
;    APPLICANT:  Kakavas, Stephan J.
;    TITLE OF INVENTION:  NOVEL POLYKETIDE DERIVATIVES
;    TITLE OF INVENTION:  AND RECOMBINANT METHODS FOR MAKING SAME
;    NUMBER OF SEQUENCES:  34
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Abbott Laboratories
;      STREET:  100 Abbott Park Rd.
;      CITY:  Abbott Park
;      STATE:  Illinois
;      COUNTRY:  USA
;      ZIP:  60064-3500
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/997,467
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/858,003
;      FILING DATE:  16-MAY-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Dianne Casuto
;      REGISTRATION NUMBER:  P-40,943
;      REFERENCE/DOCKET NUMBER:  4952.US.P2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (847)-938-3137
;      TELEFAX:  (847)-938-2623
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  29:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1010 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
US-08-997-467-29

  Query Match           35.6%;  Score 337.4;  DB 4;  Length 1010;
  Best Local Similarity   61.4%;  Pred. No. 4.5e-55;
  Matches  583;  Conservative    0;  Mismatches  351;  Indels   15;  Gaps    2;

Qy          1 GTCTTCGTTTTTCCCGGGCAGGGCCCGCAATGGCCGGGCATGGGAAGGGAACTTCTCGAC 60
              || ||||| || ||||| |||||| |||| ||| | || ||||    ||  || || ||
Db         13 GTGTTCGTGTTCCCCGGCCAGGGCTCGCAGTGGGCCGGAATGGCCGAGGGGCTGCTGGAG 72
```

```
Qy         61 GCTTCCGACGTCTTCCGGGAGAGCGTCCGCGCCTGCGAAGCCGCGTTCGCGCCCTACGTC 120
              |||| ||  ||||||       | |  | | ||||| ||||||  |   ||| ||| ||
Db         73 CGGTCCGGCGCGTTCCGGAGTGCGGCCGACTCGTGCGACGCCGCGCTGCGGCCGTACCTC 132

Qy        121 GACTGGTCGGTGGAGCAGGTGTTGCGGGACTCGCCGGACGCTCCCGGGCTGGACCGGGTG 180
              | ||||||||||  |   ||| |||| |     |||||||| |||  ||| ||||||||
Db        133 GGCTGGTCGGTGCTGAGCGTGCTGCGCGGGGAACCGGACGCGCCCTCGCTCGACCGGGTC 192

Qy        181 GACGTCGTCCAGCCGACCCTGTTCGCCGTCATGATCTCCCTGGCCGCCCTCTGGCGCTCG 240
              ||||||||  ||||||   ||||||  |  | ||| |||| || || ||  |||||||  ||
Db        193 GACGTCGTGCAGCCGGTGCTGTTCACGATGATGGTCTCGCTCGCGGCGGTCTGGCGTGCG 252

Qy        241 CAAGGGGTCGAGCCGTGCGCGGTGCTGGGACACAGCCTGGGCGAGATCGCGGCAGCCCAC 300
              |  |||||  || |||   |||||  | || |||   | ||| ||||||||  || |||||
Db        253 CTGGGGGTGGAACCGGCGGCGGTCGTCGGGCACTCGCAGGGTGAGATCGCCGCTGCCCAT 312

Qy        301 GTCTCGGGAGGCCTGTCCCTGGCCGACGCCGCACGCGTGGTGACGCTTTGGAGCCAGGCA 360
              ||| | || |  |||| |||| |||| | || ||  | ||  | ||  | || | |||
Db        313 GTCGCCGGTGCGCTGTCGCTGGACGACTCGGCCCGGATCGTCGCCCTGCGCAGTCGGGCG 372

Qy        361 CAGACCACCCTTGCCGGGACCGGCGCGCTCGTCTCCGTCGCCGCCACGCCGGATGAGCTC 420
              |  |   || || || |  ||||   | ||  | ||  |     || |||| |||||
Db        373 TGGCTCGGACTGGCGGGCAAGGGCGGCATGGTGGCGGTGCCGATGCCGGCGGAGGAGCTG 432

Qy        421 CTGCCCCGAATCGCTCCGTGGACCGAGGACAACCCGGCGCGGCTCGCCGTCGCAGCCGTC 480
              | ||| ||  | |   |||||              ||  || || |||||||| ||||||
Db        433 CGGCCGCGGCTGGTGACGTGG------------GGGGACCGTCTGGCCGTCGCCGCCGTC 480

Qy        481 AACGGACCCCGGAGCACAGTCGTTTCCGGTGCCCGCGAGGCCGTCGCGGACCTGGTGGCC 540
              ||| | ||| |   |   | |||  | || | ||  |||||  | || || ||||||||
Db        481 AACAGCCCCGGTTCCTGCGCCGTCGCAGGCGACCCGGAGGCGCTGGCCGAACTGGTGGCG 540

Qy        541 GACCTCACCGCCGCGCAGGTGCGCACGCGCATGATCCCGGTGGACGTTCCCGC---CCAC 597
              || ||||  | |  ||||| | | ||   ||||||  |  | ||   | ||   ||||
Db        541 CTGCTGACCGGTGAGGGGGTGCACGCCCGGCCGATCCCCGGCGTCGACACGGCGGGCCAC 600

Qy        598 TCCCCCCTGATGTACGCCATCGAGGAACGGGTCGTCAGCGGCCTGCTGCCCATCACCCCA 657
              || || | | || ||||  |   ||  |   |  |     |  |||   ||  || ||||
Db        601 TCGCCGCAGGTGGACGCGTTGCGGGCTCATCTGCTGGAGGTGCTGGCCCCGGTCGCCCCC 660

Qy        658 CGCCCCTCCCGCATCCCCTTCCACTCCTCGGTGACCGGCGGCCGCCTCGACACCCGCGAG 717
              || ||  ||  |||||| ||| ||||  ||||||||||||  |  || |||  |  ||||
Db        661 CGACCGGCCGACATCCCGTTCTACTCGACGGTGACCGGCGGGCTGCTGGACGGCACCGAG 720

Qy        718 CTAGACGCGGCGTACTGGTACCGCAACATGTCGAGCACGGTCCGGTTCGAGCCCGCCGCC 777
              || |||||| ||||||||||||||||||        | |||  ||||||||  ||| |
Db        721 CTGGACGCGACGTACTGGTACCGCAACATGCGCGAGCCCGTCGAGTTCGAGCGGGCCACA 780

Qy        778 CGGCTGCTTCTGCAGCAGGGGCCCAAGACGTTCGTCGAGATGAGCCCGCACCCGGTGCTG 837
              |||  |||  |      | |||| | |    ||| | |||| |||||||||  ||  |||||
Db        781 CGGGCGCTGATCGCCGACGGGCACGACGTCTTCCTGGAGACGAGCCCGCATCCCATGCTG 840

Qy        838 ACCATGGGCCTCCAGGAGCTCGCCCCGGACCTGGGCGACACCACCGGCACCGCCGACACC 897
              || |||  ||  || ||  | | | |||   |||  |  | | |       ||  |||
Db        841 GCCGTGGCGCTGGAGCAGACGGTCACCGACGCCGGCACCGACGCGGCGGTGCTCGGGACC 900

Qy        898 GTGATCATGGGCACGCTGCGCCGCGGCCAGGGCACCCTGGACCACTTCC 946
              ||  |    ||       | |   ||  |    |||||| |  || ||
Db        901 CTGCGCCGCCGCCACGGCGGTCCTCGCGCGCTGGCCCTGGCCGTCTGCC 949
```

Search completed: June 18, 2003, 01:06:02
Job time : 57.0664 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 17, 2003, 17:56:34 ; Search time 112.033 Seconds
(without alignments)
12253.192 Million cell updates/sec

Title: US-09-914-286-1_COPY_85_1032
Perfect score: 948
Sequence: 1 gtcttcgtttttcccgggca.....gcaccctggaccacttcctg 948

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1029858 seqs, 724030393 residues

Total number of hits satisfying chosen parameters: 2059716

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
11: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
12: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
13: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
14: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	337.4	35.6	1010	9	US-09-735-056-29	Sequence 29, Appl
2	284	30.0	13842	9	US-09-860-846-30	Sequence 30, Appl
3	284	30.0	13842	9	US-09-988-384B-30	Sequence 30, Appl
4	284	30.0	13842	9	US-09-836-821-30	Sequence 30, Appl
5	284	30.0	13842	10	US-09-861-289-30	Sequence 30, Appl
6	284	30.0	36778	9	US-09-860-846-5	Sequence 5, Appli
7	284	30.0	36778	9	US-09-836-821-5	Sequence 5, Appli
8	284	30.0	36778	10	US-09-861-289-5	Sequence 5, Appli
9	284	30.0	37948	9	US-09-988-384B-5	Sequence 5, Appli
10	284	30.0	38506	9	US-09-793-708-19	Sequence 19, Appl
11	279.6	29.5	1035	9	US-09-735-056-30	Sequence 30, Appl
12	267.2	28.2	11220	9	US-09-860-846-32	Sequence 32, Appl
13	267.2	28.2	11220	9	US-09-988-384B-32	Sequence 32, Appl
14	267.2	28.2	11220	9	US-09-836-821-32	Sequence 32, Appl
15	267.2	28.2	11220	10	US-09-861-289-32	Sequence 32, Appl
16	262.6	27.7	4689	9	US-09-860-846-34	Sequence 34, Appl
17	262.6	27.7	4689	9	US-09-988-384B-34	Sequence 34, Appl
18	262.6	27.7	4689	9	US-09-836-821-34	Sequence 34, Appl
19	262.6	27.7	4689	10	US-09-861-289-34	Sequence 34, Appl
20	253.2	26.7	50937	9	US-09-808-880-1	Sequence 1, Appli
21	250.6	26.4	4041	9	US-09-860-846-36	Sequence 36, Appl
22	250.6	26.4	4041	9	US-09-988-384B-36	Sequence 36, Appl
23	250.6	26.4	4041	9	US-09-836-821-36	Sequence 36, Appl
24	250.6	26.4	4041	10	US-09-861-289-36	Sequence 36, Appl
25	231	24.4	68750	9	US-10-014-717-1	Sequence 1, Appli
26	200.4	21.1	15872	9	US-09-860-846-1	Sequence 1, Appli
27	200.4	21.1	15872	9	US-09-988-384B-1	Sequence 1, Appli
28	200.4	21.1	15872	9	US-09-836-821-1	Sequence 1, Appli
29	200.4	21.1	15872	10	US-09-861-289-1	Sequence 1, Appli
30	143.4	15.1	426	9	US-10-125-815-6	Sequence 6, Appli
31	134.4	14.2	1030	9	US-09-735-056-2	Sequence 2, Appli
32	119.4	12.6	426	9	US-10-125-815-5	Sequence 5, Appli
33	103.8	10.9	925	9	US-09-735-056-1	Sequence 1, Appli
34	102.6	10.8	4851	9	US-09-712-363-116	Sequence 116, App
35	92.4	9.7	4209	9	US-09-712-363-20	Sequence 20, Appl
36	79.2	8.4	5484	9	US-09-712-363-115	Sequence 115, App
37	68.8	7.3	2561	9	US-09-976-740-48	Sequence 48, Appl
38	68.8	7.3	2561	12	US-10-023-529-48	Sequence 48, Appl
39	68.8	7.3	2561	12	US-10-023-523-48	Sequence 48, Appl
40	60.6	6.4	1293	9	US-10-145-415-31	Sequence 31, Appl
41	59.6	6.3	1161	10	US-09-808-387-5	Sequence 5, Appli
42	59.6	6.3	1674	10	US-09-808-387-3	Sequence 3, Appli
43	59.6	6.3	1725	10	US-09-808-387-1	Sequence 1, Appli
c 44	59.2	6.2	1771	9	US-09-991-496-7	Sequence 7, Appli
c 45	59.2	6.2	1771	10	US-09-874-923-7	Sequence 7, Appli

ALIGNMENTS

RESULT 1
US-09-735-056-29
; Sequence 29, Application US/09735056
; Publication No. US20030013662A1
; GENERAL INFORMATION:
; APPLICANT: Katz, Leonard
; APPLICANT: Stassi, Diane L.
; APPLICANT: Summers Jr., Richard G.
; APPLICANT: Ruan, Xiaoan
; APPLICANT: Pereda-Lopez, Ana
; APPLICANT: Kakavas, Stephan J.
; TITLE OF INVENTION: NOVEL POLYKETIDE DERIVATIVES
; TITLE OF INVENTION: AND RECOMBINANT METHODS FOR MAKING SAME
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Abbott Laboratories
; STREET: 100 Abbott Park Rd.
; CITY: Abbott Park
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60064-3500
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/735,056
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/858,003
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Dianne Casuto
; REGISTRATION NUMBER: P-40,943
; REFERENCE/DOCKET NUMBER: 4952.US.P2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (847)-938-3137
; TELEFAX: (847)-938-2623
; TELEX:
; INFORMATION FOR SEQ ID NO: 29:

```
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1010 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
US-09-735-056-29

  Query Match             35.6%;  Score 337.4;  DB 9;  Length 1010;
  Best Local Similarity   61.4%;  Pred. No. 1.2e-68;
  Matches  583;  Conservative    0;  Mismatches  351;  Indels   15;  Gaps    2;

Qy          1 GTCTTCGTTTTTCCCGGGCAGGGCCCGCAATGGCCGGGCATGGGAAGGGAACTTCTCGAC 60
              || ||||| || ||||| |||||| |||| ||| | || ||||    ||  || || ||
Db         13 GTGTTCGTGTTCCCCGGCCAGGGCTCGCAGTGGGCCGGAATGGCCGAGGGGCTGCTGGAG 72

Qy         61 GCTTCCGACGTCTTCCGGGAGAGCGTCCGCGCCTGCGAAGCCGCGTTCGCGCCCTACGTC 120
                 |||| ||  ||||||      | |  | | ||||| ||||||  |    ||| ||| ||
Db         73 CGGTCCGGCGCGTTCCGGAGTGCGGCCGACTCGTGCGACGCCGCGCTGCGGCCGTACCTC 132

Qy        121 GACTGGTCGGTGGAGCAGGTGTTGCGGGACTCGCCGGACGCTCCCGGGCTGGACCGGGTG 180
              | |||||||||| |   ||| |||| |     |||||||| |||  ||| ||||||||
Db        133 GGCTGGTCGGTGCTGAGCGTGCTGCGCGGGGAACCGGACGCGCCCTCGCTCGACCGGGTC 192

Qy        181 GACGTCGTCCAGCCGACCCTGTTCGCCGTCATGATCTCCCTGGCCGCCCTCTGGCGCTCG 240
              |||||||| ||||||   |||||| |  | ||| |||| || || ||  ||||||| ||
Db        193 GACGTCGTGCAGCCGGTGCTGTTCACGATGATGGTCTCGCTCGCGGCGGTCTGGCGTGCG 252

Qy        241 CAAGGGGTCGAGCCGTGCGCGGTGCTGGGACACAGCCTGGGCGAGATCGCGGCAGCCCAC 300
              |  ||||| || |||   |||||  | || |||   | ||| ||||||| || ||||||
Db        253 CTGGGGGTGGAACCGGCGGCGGTCGTCGGGCACTCGCAGGGTGAGATCGCCGCTGCCCAT 312

Qy        301 GTCTCGGGAGGCCTGTCCCTGGCCGACGCCGCACGCGTGGTGACGCTTTGGAGCCAGGCA 360
              ||| | || |  |||| |||| |||| | || ||  | ||  | ||  | || | |||
Db        313 GTCGCCGGTGCGCTGTCGCTGGACGACTCGGCCCGGATCGTCGCCCTGCGCAGTCGGGCG 372

Qy        361 CAGACCACCCTTGCCGGGACCGGCGCGCTCGTCTCCGTCGCCGCCACGCCGGATGAGCTC 420
                |  |   || || || |  ||||   | ||  | ||  |      || |||| |||||
Db        373 TGGCTCGGACTGGCGGGCAAGGGCGGCATGGTGGCGGTGCCGATGCCGGCGGAGGAGCTG 432

Qy        421 CTGCCCCGAATCGCTCCGTGGACCGAGGACAACCCGGCGCGGCTCGCCGTCGCAGCCGTC 480
              | ||| ||  | |  |||||                ||  || || |||||||| ||||||
Db        433 CGGCCGCGGCTGGTGACGTGG------------GGGGACCGTCTGGCCGTCGCCGCCGTC 480

Qy        481 AACGGACCCCGGAGCACAGTCGTTTCCGGTGCCCGCGAGGCCGTCGCGGACCTGGTGGCC 540
              ||| | ||| |   |   | |||  | || | ||  |||||  | || || ||||||||
Db        481 AACAGCCCCGGTTCCTGCGCCGTCGCAGGCGACCCGGAGGCGCTGGCCGAACTGGTGGCG 540

Qy        541 GACCTCACCGCCGCGCAGGTGCGCACGCGCATGATCCCGGTGGACGTTCCCGC---CCAC 597
                || ||||  | |  ||||| | | ||   |||||| |  | ||   | ||   ||||
Db        541 CTGCTGACCGGTGAGGGGGTGCACGCCCGGCCGATCCCCGGCGTCGACACGGCGGGCCAC 600

Qy        598 TCCCCCCTGATGTACGCCATCGAGGAACGGGTCGTCAGCGGCCTGCTGCCCATCACCCCA 657
              || || | | || ||||   |   ||  |   |  |     ||   || ||||     |
Db        601 TCGCCGCAGGTGGACGCGTTGCGGGCTCATCTGCTGGAGGTGCTGGCCCCGGTCGCCCCC 660

Qy        658 CGCCCCTCCCGCATCCCCTTCCACTCCTCGGTGACCGGCGGCCGCCTCGACACCCGCGAG 717
              || ||  ||  |||||| ||| ||||  |||||||||||||  |  || |||  |  ||||
Db        661 CGACCGGCCGACATCCCGTTCTACTCGACGGTGACCGGCGGGCTGCTGGACGGCACCGAG 720

Qy        718 CTAGACGCGGCGTACTGGTACCGCAACATGTCGAGCACGGTCCGGTTCGAGCCCGCCGCC 777
              || |||||| |||||||||||||||||||        | |||  ||||||||  ||| |
Db        721 CTGGACGCGACGTACTGGTACCGCAACATGCGCGAGCCCGTCGAGTTCGAGCGGGCCACA 780

Qy        778 CGGCTGCTTCTGCAGCAGGGGCCCAAGACGTTCGTCGAGATGAGCCCGCACCCGGTGCTG 837
              |||  |||  |     | |||| | |     ||| | |||| |||||||||  ||  |||||
Db        781 CGGGCGCTGATCGCCGACGGGCACGACGTCTTCCTGGAGACGAGCCCGCATCCCATGCTG 840

Qy        838 ACCATGGGCCTCCAGGAGCTCGCCCCGGACCTGGGCGACACCACCGGCACCGCCGACACC 897
              || |||  ||  || ||   | | | |||  |||  |  | | |         ||  |||
Db        841 GCCGTGGCGCTGGAGCAGACGGTCACCGACGCCGGCACCGACGCGGCGGTGCTCGGGACC 900

Qy        898 GTGATCATGGGCACGCTGCGCCGCGGCCAGGGCACCCTGGACCACTTCC 946
                ||  |    ||       | |  ||  |    |||||| |  || ||

Db        901 CTGCGCCGCCGCCACGGCGGTCCTCGCGCGCTGGCCCTGGCCGTCTGCC 949


RESULT 2
US-09-860-846-30
; Sequence 30, Application US/09860846
; Patent No. US20020164742A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/860,846
;  CURRENT FILING DATE:  2001-05-18
;  PRIOR APPLICATION NUMBER: 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 30
;   LENGTH: 13842
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-860-846-30

  Query Match             30.0%;  Score 284;  DB 9;  Length 13842;
  Best Local Similarity   60.5%;  Pred. No. 1.3e-56;
  Matches  530;  Conservative    0;  Mismatches  325;  Indels   21;  Gaps    3;

Qy          4 TTCGTTTTTCCCGGGCAGGGCCCGCAATGGCCGGGCATGGGAAGGGAACTTCTCGACGCT 63
              ||||| || ||||| |||||| |||| ||| | |||||||    ||||| || ||| ||
Db       4948 TTCGTGTTCCCCGGTCAGGGCACGCAGTGGGCTGGCATGGGTGCCGAACTGCTGGACTCT 5007

Qy         64 TCCGACGTCTTCCGGGAGAGCGTCCGCGCCTGCGAAGCCGCGTTCGCGCCCTACGTCGAC 123
              ||||  || |||  || |  | |   ||  || || |||||| |  | || |||||||||
Db       5008 TCCGCGGTGTTCGCGGCGGCCATGGCCGAGTGTGAGGCCGCGCTGTCCCCGTACGTCGAC 5067

Qy        124 TGGTCGGTGGAGCAGGTGTTGCGGGACTCGCCGGACGCTCCCGGGCTGGACCGGGTGGAC 183
              |||||| |||||   ||  | ||| |  | || |  || |||  |||||| ||||| ||
Db       5068 TGGTCGCTGGAGGCCGTCGTACGGCAGGCCCCCGGTGCGCCCACGCTGGAGCGGGTCGAT 5127

Qy        184 GTCGTCCAGCCGACCCTGTTCGCCGTCATGATCTCCCTGGCCGCCCTCTGGCGCTCGCAA 243
              ||||| |||||       |||||||||||| |||| |||||    | | ||||      ||
Db       5128 GTCGTGCAGCCTGTGACGTTCGCCGTCATGGTCTCGCTGGCTCGCGTGTGGCAGCACCAC 5187

Qy        244 GGGGTCGAGCCGTGCGCGGTGCTGGGACACAGCCTGGGCGAGATCGCGGCAGCCCACGTC 303
              || ||   |||    |||||  | || |||   | ||||||||||| || ||  |||||
Db       5188 GGTGTGACGCCCCAGGCGGTCGTCGGCCACTCGCAGGGCGAGATCGCCGCCGCGTACGTC 5247

Qy        304 TCGGGAGGCCTGTCCCTGGCCGACGCCGCACGCGTGGTGACGCTTTGGAGCCAGGCACAG 363
              | |||| |||| |||||| |||||||| ||||| || || ||  | ||| || |
Db       5248 GCCGGAGCCCTGCCCCTGGACGACGCCGCCCGCGTCGTCACCCTGCGCAGCAAGTCCATC 5307

Qy        364 ACCAC---CCTTGCCGGGACCGGCGCGCTCGTCTCCGTCGCCGCCACGCCGGATGAGCTC 420
              || |   ||| ||||| |  ||||    |  | ||| ||||    |   ||| |   ||
Db       5308 GCCGCCCACCTCGCCGGCAAGGGCGGCATGCTGTCCCTCGCGCTGAACGAGGACGCCGTC 5367

Qy        421 CTGCCCCGAATCGCTCCGTGGACCGAGGACAACCCGGCGCGGCTCGCCGTCGCAGCCGTC 480
              |||   ||| |              |    ||   |  ||||  |||||| ||||||
Db       5368 CTGGAGCGACT--------------GAGTGACTTCGACGGGCTGTCCGTCGCCGCCGTC 5412

Qy        481 AACGGACCCCGGAGCACAGTCGTTTCCGGTGCCCGCGAGGCCGTCGCGGACCTGGTGGCC 540
              ||||| |||     ||| ||||| || |||| || ||      |||  || || |
Db       5413 AACGGGCCCACCGCCACTGTCGTGTCGGGTGACCCCGTACAGATCGAAGAGCTTGCTCAG 5472

Qy        541 GACCTCACCGCCGCGCAGGTGCGCACGCGCATGATCCCGGTGGACGTTCCCGCCCACTCC 600
              |     || || |       | ||| |||| || || || || |||     | |||||  |
Db       5473 GCGTGCAAGGCGGACGGATTCCGCGCGCGGATCATTCCCGTCGACTACGCGTCCCACAGC 5532

Qy        601 CCCCTGATGTACGCCATCGAGGAACGGGTCGTCAGCGGCCTGCTGCCCATCACCCCACGC 660
              | | | |  |  |||||||     | ||| |   | |||         ||| ||| |
Db       5533 CGGCAGGTCGAGATCATCGAGAGCGAGCTCGCCCAGGTCCTCGCCGGTCTCAGCCCGCAG 5592
```

```
Qy       661 CCCTCCCGCATCCCCTTCCACTCCTCGGTGACCGGCGGCCGCCTCGACACCCGCGAGCTA 720
Db      5593 GCCCCGCGCGTGCCGTTCTTCTCGACGCTCGAAGGCACCTGGATCACCGAGCCCGTCCTC 5652

Qy       721 GACGCGGCGTACTGGTACCGCAACATGTCGAGCACGGTCCGGTTCGAGCCCGCCGCCCGG 780
Db      5653 GACGGCACCTACTGGTACCGCAACCTCCGTCACCGCGTCGGCTTCGCCCCCGCCATCGAG 5712

Qy       781 ---CTGCTTCTGCAGCAGGGGCCCAAGACGTTCGTCGAGATGAGCCCGCACCCGGTGCTG 837
Db      5713 ACCCTGGCCGTCGACGAGGGCTTCACGCACTTCGTCGAGGTCAGCGCCCACCCCGTCCTC 5772

Qy       838 ACCATGGGCCTCCAGGAGCTCGCCCCGGACCTGGGC 873
Db      5773 ACCATGACCCTCCCCGAGACCGTCACCGGCCTCGGC 5808


RESULT 3
US-09-988-384B-30
; Sequence 30, Application US/09988384B
; Publication No. US20030073824A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.536US1
;  CURRENT APPLICATION NUMBER: US/09/988,384B
;  CURRENT FILING DATE:  2001-11-19
;  PRIOR APPLICATION NUMBER: PCT/US99/14398
;  PRIOR FILING DATE: 1999-06-25
;  PRIOR APPLICATION NUMBER: US 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 53
; SEQ ID NO 30
;   LENGTH: 13842
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-988-384B-30

  Query Match             30.0%;  Score 284;  DB 9;  Length 13842;
  Best Local Similarity   60.5%;  Pred. No. 1.3e-56;
  Matches  530;  Conservative    0;  Mismatches  325;  Indels   21;  Gaps    3;

Qy         4 TTCGTTTTTCCCGGGCAGGGCCCGCAATGGCCGGGCATGGGAAGGGAACTTCTCGACGCT 63
Db      4948 TTCGTGTTCCCCGGTCAGGGCACGCAGTGGGCTGGCATGGGTGCCGAACTGCTGGACTCT 5007

Qy        64 TCCGACGTCTTCCGGGAGAGCGTCCGCGCCTGCGAAGCCGCGTTCGCGCCCTACGTCGAC 123
Db      5008 TCCGCGGTGTTCGCGGCGGCCATGGCCGAGTGTGAGGCCGCGCTGTCCCCGTACGTCGAC 5067

Qy       124 TGGTCGGTGGAGCAGGTGTTGCGGGACTCGCCGGACGCTCCCGGGCTGGACCGGGTGGAC 183
Db      5068 TGGTCGCTGGAGGCCGTCGTACGGCAGGCCCCCGGTGCGCCCACGCTGGAGCGGGTCGAT 5127

Qy       184 GTCGTCCAGCCGACCCTGTTCGCCGTCATGATCTCCCTGGCCGCCCTCTGGCGCTCGCAA 243
Db      5128 GTCGTGCAGCCTGTGACGTTCGCCGTCATGGTCTCGCTGGCTCGCGTGTGGCAGCACCAC 5187

Qy       244 GGGGTCGAGCCGTGCGCGGTGCTGGGACACAGCCTGGGCGAGATCGCGGCAGCCCACGTC 303
Db      5188 GGTGTGACGCCCCAGGCGGTCGTCGGCCACTCGCAGGGCGAGATCGCCGCCGCGTACGTC 5247

Qy       304 TCGGGAGGCCTGTCCCTGGCCGACGCCGCACGCGTGGTGACGCTTTGGAGCCAGGCACAG 363
Db      5248 GCCGGAGCCCTGCCCCTGGACGACGCCGCCCGCGTCGTCACCCTGCGCAGCAAGTCCATC 5307

Qy       364 ACCAC---CCTTGCCGGGACCGGCGCGCTCGTCTCCGTCGCCGCCACGCCGGATGAGCTC 420
Db      5308 GCCGCCCACCTCGCCGGCAAGGGCGGCATGCTGTCCCTCGCGCTGAACGAGGACGCCGTC 5367

Qy       421 CTGCCCCGAATCGCTCCGTGGACCGAGGACAACCCGGCGCGGCTCGCCGTCGCAGCCGTC 480
Db      5368 CTGGAGCGACT---------------GAGTGACTTCGACGGGCTGTCCGTCGCCGCCGTC 5412

Qy       481 AACGGACCCCGGAGCACAGTCGTTTCCGGTGCCCGCGAGGCCGTCGCGGACCTGGTGGCC 540
Db      5413 AACGGGCCCACCGCCACTGTCGTGTCGGGTGACCCCGTACAGATCGAAGAGCTTGCTCAG 5472

Qy       541 GACCTCACCGCCGCGCAGGTGCGCACGCGCATGATCCCGGTGGACGTTCCCGCCCACTCC 600
Db      5473 GCGTGCAAGGCGGACGGATTCCGCGCGCGGATCATTCCCGTCGACTACGCGTCCCACAGC 5532

Qy       601 CCCCTGATGTACGCCATCGAGGAACGGGTCGTCAGCGGCCTGCTGCCCATCACCCCACGC 660
Db      5533 CGGCAGGTCGAGATCATCGAGAGCGAGCTCGCCCAGGTCCTCGCCGGTCTCAGCCCGCAG 5592

Qy       661 CCCTCCCGCATCCCCTTCCACTCCTCGGTGACCGGCGGCCGCCTCGACACCCGCGAGCTA 720
Db      5593 GCCCCGCGCGTGCCGTTCTTCTCGACGCTCGAAGGCACCTGGATCACCGAGCCCGTCCTC 5652

Qy       721 GACGCGGCGTACTGGTACCGCAACATGTCGAGCACGGTCCGGTTCGAGCCCGCCGCCCGG 780
Db      5653 GACGGCACCTACTGGTACCGCAACCTCCGTCACCGCGTCGGCTTCGCCCCCGCCATCGAG 5712

Qy       781 ---CTGCTTCTGCAGCAGGGGCCCAAGACGTTCGTCGAGATGAGCCCGCACCCGGTGCTG 837
Db      5713 ACCCTGGCCGTCGACGAGGGCTTCACGCACTTCGTCGAGGTCAGCGCCCACCCCGTCCTC 5772

Qy       838 ACCATGGGCCTCCAGGAGCTCGCCCCGGACCTGGGC 873
Db      5773 ACCATGACCCTCCCCGAGACCGTCACCGGCCTCGGC 5808


RESULT 4
US-09-836-821-30
; Sequence 30, Application US/09836821
; Publication No. US20030087405A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/836,821
;  CURRENT FILING DATE:  2001-04-17
;  PRIOR APPLICATION NUMBER: 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 30
;   LENGTH: 13842
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-836-821-30

  Query Match             30.0%;  Score 284;  DB 9;  Length 13842;
  Best Local Similarity   60.5%;  Pred. No. 1.3e-56;
  Matches  530;  Conservative    0;  Mismatches  325;  Indels   21;  Gaps    3;

Qy         4 TTCGTTTTTCCCGGGCAGGGCCCGCAATGGCCGGGCATGGGAAGGGAACTTCTCGACGCT 63
Db      4948 TTCGTGTTCCCCGGTCAGGGCACGCAGTGGGCTGGCATGGGTGCCGAACTGCTGGACTCT 5007

Qy        64 TCCGACGTCTTCCGGGAGAGCGTCCGCGCCTGCGAAGCCGCGTTCGCGCCCTACGTCGAC 123
Db      5008 TCCGCGGTGTTCGCGGCGGCCATGGCCGAGTGTGAGGCCGCGCTGTCCCCGTACGTCGAC 5067

Qy       124 TGGTCGGTGGAGCAGGTGTTGCGGGACTCGCCGGACGCTCCCGGGCTGGACCGGGTGGAC 183
Db      5068 TGGTCGCTGGAGGCCGTCGTACGGCAGGCCCCCGGTGCGCCCACGCTGGAGCGGGTCGAT 5127
```

```
Qy        184 GTCGTCCAGCCGACCCTGTTCGCCGTCATGATCTCCCTGGCCGCCCTCTGGCGCTCGCAA 243
              ||||| |||||       |||||||||||||| |||| |||||    | | ||||      ||
Db       5128 GTCGTGCAGCCTGTGACGTTCGCCGTCATGGTCTCGCTGGCTCGCGTGTGGCAGCACCAC 5187

Qy        244 GGGGTCGAGCCGTGCGCGGTGCTGGGACACAGCCTGGGCGAGATCGCGGCAGCCCACGTC 303
              || ||   |||    |||||  | || |||   | ||||||||||| || ||  |||||
Db       5188 GGTGTGACGCCCCAGGCGGTCGTCGGCCACTCGCAGGGCGAGATCGCCGCCGCGTACGTC 5247

Qy        304 TCGGGAGGCCTGTCCCTGGCCGACGCCGCACGCGTGGTGACGCTTTGGAGCCAGGCACAG 363
              | |||| |||| |||||| ||||||||| ||||| || || ||  | ||| || |
Db       5248 GCCGGAGCCCTGCCCCTGGACGACGCCGCCCGCGTCGTCACCCTGCGCAGCAAGTCCATC 5307

Qy        364 ACCAC---CCTTGCCGGGACCGGCGCGCTCGTCTCCGTCGCCGCCACGCCGGATGAGCTC 420
              || |   ||| ||||| |  ||||    |  | ||| ||||     |    ||| |   ||
Db       5308 GCCGCCCACCTCGCCGGCAAGGGCGGCATGCTGTCCCTCGCGCTGAACGAGGACGCCGTC 5367

Qy        421 CTGCCCCGAATCGCTCCGTGGACCGAGGACAACCCGGCGCGGCTCGCCGTCGCAGCCGTC 480
              |||   ||| |                |    ||   |   ||||  ||||||| ||||||
Db       5368 CTGGAGCGACT---------------GAGTGACTTCGACGGGCTGTCCGTCGCCGCCGTC 5412

Qy        481 AACGGACCCCGGAGCACAGTCGTTTCCGGTGCCCGCGAGGCCGTCGCGGACCTGGTGGCC 540
              ||||| |||     ||| ||||| || |||| || ||      |||  || || |
Db       5413 AACGGGCCCACCGCCACTGTCGTGTCGGGTGACCCCGTACAGATCGAAGAGCTTGCTCAG 5472

Qy        541 GACCTCACCGCCGCGCAGGTGCGCACGCGCATGATCCCGGTGGACGTTCCCGCCCACTCC 600
              |    ||  || |       | ||| |||| || || || || |||    | |||||  |
Db       5473 GCGTGCAAGGCGGACGGATTCCGCGCGCGGATCATTCCCGTCGACTACGCGTCCCACAGC 5532

Qy        601 CCCCTGATGTACGCCATCGAGGAACGGGTCGTCAGCGGCCTGCTGCCCATCACCCCACGC 660
              |  | | |  |  |||||||    | ||| |   | |||         ||| ||| |
Db       5533 CGGCAGGTCGAGATCATCGAGAGCGAGCTCGCCCAGGTCCTCGCCGGTCTCAGCCCGCAG 5592

Qy        661 CCCTCCCGCATCCCCTTCCACTCCTCGGTGACCGGCGGCCGCCTCGACACCCGCGAGCTA 720
              || | ||| | || |||  |||  || |    |||  | |  ||  |   | ||  ||
Db       5593 GCCCCGCGCGTGCCGTTCTTCTCGACGCTCGAAGGCACCTGGATCACCGAGCCCGTCCTC 5652

Qy        721 GACGCGGCGTACTGGTACCGCAACATGTCGAGCACGGTCCGGTTCGAGCCCGCCGCCCGG 780
              ||||   | ||||||||||||||| |       |  ||| | ||||  ||||||  | |
Db       5653 GACGGCACCTACTGGTACCGCAACCTCCGTCACCGCGTCGGCTTCGCCCCCGCCATCGAG 5712

Qy        781 ---CTGCTTCTGCAGCAGGGGCCCAAGACGTTCGTCGAGATGAGCCCGCACCCGGTGCTG 837
                 |||    |  |  ||||   || |   ||||||||| | ||| | ||||| || ||
Db       5713 ACCCTGGCCGTCGACGAGGGCTTCACGCACTTCGTCGAGGTCAGCGCCCACCCCGTCCTC 5772

Qy        838 ACCATGGGCCTCCAGGAGCTCGCCCCGGACCTGGGC 873
              ||||||  |||||  |||  || | | | ||| |||
Db       5773 ACCATGACCCTCCCCGAGACCGTCACCGGCCTCGGC 5808


RESULT 5
US-09-861-289-30
; Sequence 30, Application US/09861289
; Patent No. US20020110897A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/861,289
;  CURRENT FILING DATE:  2001-05-18
;  PRIOR APPLICATION NUMBER: 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 30
;   LENGTH: 13842
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-861-289-30

  Query Match             30.0%;  Score 284;  DB 10;  Length 13842;
  Best Local Similarity   60.5%;  Pred. No. 1.3e-56;
  Matches  530;  Conservative    0;  Mismatches  325;  Indels   21;  Gaps    3;

Qy          4 TTCGTTTTTCCCGGGCAGGGCCCGCAATGGCCGGGCATGGGAAGGGAACTTCTCGACGCT 63
              ||||| || ||||| |||||| |||| ||| | ||||||||     ||||| || ||| ||
Db       4948 TTCGTGTTCCCCGGTCAGGGCACGCAGTGGGCTGGCATGGGTGCCGAACTGCTGGACTCT 5007

Qy         64 TCCGACGTCTTCCGGGAGAGCGTCCGCGCCTGCGAAGCCGCGTTCGCGCCCTACGTCGAC 123
              ||||  || |||  || |  | |    ||  || || ||||||  |  | || |||||||||
Db       5008 TCCGCGGTGTTCGCGGCGGCCATGGCCGAGTGTGAGGCCGCGCTGTCCCCGTACGTCGAC 5067

Qy        124 TGGTCGGTGGAGCAGGTGTTGCGGGACTCGCCGGACGCTCCCGGGCTGGACCGGGTGGAC 183
              |||||| |||||    ||  | ||| |  | || |  || |||  |||||| ||||| ||
Db       5068 TGGTCGCTGGAGGCCGTCGTACGGCAGGCCCCCGGTGCGCCCACGCTGGAGCGGGTCGAT 5127

Qy        184 GTCGTCCAGCCGACCCTGTTCGCCGTCATGATCTCCCTGGCCGCCCTCTGGCGCTCGCAA 243
              ||||| |||||       |||||||||||||| |||| |||||    | | ||||      ||
Db       5128 GTCGTGCAGCCTGTGACGTTCGCCGTCATGGTCTCGCTGGCTCGCGTGTGGCAGCACCAC 5187

Qy        244 GGGGTCGAGCCGTGCGCGGTGCTGGGACACAGCCTGGGCGAGATCGCGGCAGCCCACGTC 303
              || ||   |||    |||||  | || |||   | ||||||||||| || ||  |||||
Db       5188 GGTGTGACGCCCCAGGCGGTCGTCGGCCACTCGCAGGGCGAGATCGCCGCCGCGTACGTC 5247

Qy        304 TCGGGAGGCCTGTCCCTGGCCGACGCCGCACGCGTGGTGACGCTTTGGAGCCAGGCACAG 363
              | |||| |||| |||||| ||||||||| ||||| || || ||  | ||| || |
Db       5248 GCCGGAGCCCTGCCCCTGGACGACGCCGCCCGCGTCGTCACCCTGCGCAGCAAGTCCATC 5307

Qy        364 ACCAC---CCTTGCCGGGACCGGCGCGCTCGTCTCCGTCGCCGCCACGCCGGATGAGCTC 420
              || |   ||| ||||| |  ||||    |  | ||| ||||     |    ||| |   ||
Db       5308 GCCGCCCACCTCGCCGGCAAGGGCGGCATGCTGTCCCTCGCGCTGAACGAGGACGCCGTC 5367

Qy        421 CTGCCCCGAATCGCTCCGTGGACCGAGGACAACCCGGCGCGGCTCGCCGTCGCAGCCGTC 480
              |||   ||| |                |    ||   |   ||||  ||||||| ||||||
Db       5368 CTGGAGCGACT---------------GAGTGACTTCGACGGGCTGTCCGTCGCCGCCGTC 5412

Qy        481 AACGGACCCCGGAGCACAGTCGTTTCCGGTGCCCGCGAGGCCGTCGCGGACCTGGTGGCC 540
              ||||| |||     ||| ||||| || |||| || ||      |||  || || |
Db       5413 AACGGGCCCACCGCCACTGTCGTGTCGGGTGACCCCGTACAGATCGAAGAGCTTGCTCAG 5472

Qy        541 GACCTCACCGCCGCGCAGGTGCGCACGCGCATGATCCCGGTGGACGTTCCCGCCCACTCC 600
              |    ||  || |       | ||| |||| || || || || |||    | |||||  |
Db       5473 GCGTGCAAGGCGGACGGATTCCGCGCGCGGATCATTCCCGTCGACTACGCGTCCCACAGC 5532

Qy        601 CCCCTGATGTACGCCATCGAGGAACGGGTCGTCAGCGGCCTGCTGCCCATCACCCCACGC 660
              |  | | |  |  |||||||    | ||| |   | |||         ||| ||| |
Db       5533 CGGCAGGTCGAGATCATCGAGAGCGAGCTCGCCCAGGTCCTCGCCGGTCTCAGCCCGCAG 5592

Qy        661 CCCTCCCGCATCCCCTTCCACTCCTCGGTGACCGGCGGCCGCCTCGACACCCGCGAGCTA 720
              || | ||| | || |||  |||  || |    |||  | |  ||  |   | ||  ||
Db       5593 GCCCCGCGCGTGCCGTTCTTCTCGACGCTCGAAGGCACCTGGATCACCGAGCCCGTCCTC 5652

Qy        721 GACGCGGCGTACTGGTACCGCAACATGTCGAGCACGGTCCGGTTCGAGCCCGCCGCCCGG 780
              ||||   | ||||||||||||||| |       |  ||| | ||||  ||||||  | |
Db       5653 GACGGCACCTACTGGTACCGCAACCTCCGTCACCGCGTCGGCTTCGCCCCCGCCATCGAG 5712

Qy        781 ---CTGCTTCTGCAGCAGGGGCCCAAGACGTTCGTCGAGATGAGCCCGCACCCGGTGCTG 837
                 |||    |  |  ||||   || |   ||||||||| | ||| | ||||| || ||
Db       5713 ACCCTGGCCGTCGACGAGGGCTTCACGCACTTCGTCGAGGTCAGCGCCCACCCCGTCCTC 5772

Qy        838 ACCATGGGCCTCCAGGAGCTCGCCCCGGACCTGGGC 873
              ||||||  |||||  |||  || | | | ||| |||
Db       5773 ACCATGACCCTCCCCGAGACCGTCACCGGCCTCGGC 5808


RESULT 6
US-09-860-846-5
; Sequence 5, Application US/09860846
; Patent No. US20020164742A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
```

```
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/860,846
;  CURRENT FILING DATE:  2001-05-18
;  PRIOR APPLICATION NUMBER: 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 5
;   LENGTH: 36778
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-860-846-5

  Query Match             30.0%;  Score 284;  DB 9;  Length 36778;
  Best Local Similarity   60.5%;  Pred. No. 1.1e-56;
  Matches  530;  Conservative    0;  Mismatches  325;  Indels   21;  Gaps    3;

Qy          4 TTCGTTTTTCCCGGGCAGGGCCCGCAATGGCCGGGCATGGGAAGGGAACTTCTCGACGCT 63
              ||||| || ||||| |||||| |||| ||| | ||||||||     ||||| || ||| ||
Db       6689 TTCGTGTTCCCCGGTCAGGGCACGCAGTGGGCTGGCATGGGTGCCGAACTGCTGGACTCT 6748

Qy         64 TCCGACGTCTTCCGGGAGAGCGTCCGCGCCTGCGAAGCCGCGTTCGCGCCCTACGTCGAC 123
              ||||  || |||  || |  | |   ||  || || |||||| |  | || |||||||||
Db       6749 TCCGCGGTGTTCGCGGCGGCCATGGCCGAGTGTGAGGCCGCGCTGTCCCCGTACGTCGAC 6808

Qy        124 TGGTCGGTGGAGCAGGTGTTGCGGGACTCGCCGGACGCTCCCGGGCTGGACCGGGTGGAC 183
              ||||| |||||   ||  | ||| |  | || |  || |||  |||||| ||||| ||
Db       6809 TGGTCGCTGGAGGCCGTCGTACGGCAGGCCCCCGGTGCGCCCACGCTGGAGCGGGTCGAT 6868

Qy        184 GTCGTCCAGCCGACCCTGTTCGCCGTCATGATCTCCCTGGCCGCCCTCTGGCGCTCGCAA 243
              ||||| |||||      |||||||||||| |||| |||||    | | ||||      ||
Db       6869 GTCGTGCAGCCTGTGACGTTCGCCGTCATGGTCTCGCTGGCTCGCGTGTGGCAGCACCAC 6928

Qy        244 GGGGTCGAGCCGTGCGCGGTGCTGGGACACAGCCTGGGCGAGATCGCGGCAGCCCACGTC 303
              || ||   |||    |||||  | || |||   | ||||||||||| || ||  |||||
Db       6929 GGTGTGACGCCCCAGGCGGTCGTCGGCCACTCGCAGGGCGAGATCGCCGCCGCGTACGTC 6988

Qy        304 TCGGGAGGCCTGTCCCTGGCCGACGCCGCACGCGTGGTGACGCTTTGGAGCCAGGCACAG 363
              | |||| |||| ||||| |||||||| ||||| || || ||  | ||| || |
Db       6989 GCCGGAGCCCTGCCCCTGGACGACGCCGCCCGCGTCGTCACCCTGCGCAGCAAGTCCATC 7048

Qy        364 ACCAC---CCTTGCCGGGACCGGCGCGCTCGTCTCCGTCGCCGCCACGCCGGATGAGCTC 420
              || |   ||| |||| |  ||||   |  ||| |||    |   ||| |   ||
Db       7049 GCCGCCCACCTCGCCGGCAAGGGCGGCATGCTGTCCCTCGCGCTGAACGAGGACGCCGTC 7108

Qy        421 CTGCCCCGAATCGCTCCGTGGACCGAGGACAACCCGGCGCGGCTCGCCGTCGCAGCCGTC 480
              |||   ||| |                |   ||   |  ||||  ||||||| ||||||
Db       7109 CTGGAGCGACT--------------GAGTGACTTCGACGGGCTGTCCGTCGCCGCCGTC 7153

Qy        481 AACGGACCCCGGAGCACAGTCGTTTCCGGTGCCCGCGAGGCCGTCGCGGACCTGGTGGCC 540
              ||||| |||     ||| ||||| || |||| || ||     |||  || || |
Db       7154 AACGGGCCCACCGCCACTGTCGTGTCGGGTGACCCCGTACAGATCGAAGAGCTTGCTCAG 7213

Qy        541 GACCTCACCGCCGCGCAGGTGCGCACGCGCATGATCCCGGTGGACGTTCCCGCCCACTCC 600
              |    ||  || |     | ||| |||| || || || || |||    |  |||||  |
Db       7214 GCGTGCAAGGCGGACGGATTCCGCGCGCGGATCATTCCCGTCGACTACGCGTCCCACAGC 7273

Qy        601 CCCCTGATGTACGCCATCGAGGAACGGGTCGTCAGCGGCCTGCTGCCCATCACCCCACGC 660
              |  | | |  |  |||||||     | ||| |   | |||         ||| ||| |
Db       7274 CGGCAGGTCGAGATCATCGAGAGCGAGCTCGCCCAGGTCCTCGCCGGTCTCAGCCCGCAG 7333

Qy        661 CCCTCCCGCATCCCCTTCCACTCCTCGGTGACCGGCGGCCGCCTCGACACCCGCGAGCTA 720
              || | ||| | || |||  |||  || |   |||  | |  ||  |   | ||  ||
Db       7334 GCCCCGCGCGTGCCGTTCTTCTCGACGCTCGAAGGCACCTGGATCACCGAGCCCGTCCTC 7393

Qy        721 GACGCGGCGTACTGGTACCGCAACATGTCGAGCACGGTCCGGTTCGAGCCCGCCGCCCGG 780
              ||||   |  |||||||||||||| |     |||  | |||||   ||  |   |  |
Db       7394 GACGGCACCTACTGGTACCGCAACCTCCGTCACCGCGTCGGCTTCGCCCCCGCCATCGAG 7453

Qy        781 ---CTGCTTCTGCAGCAGGGGCCCAAGACGTTCGTCGAGATGAGCCCGCACCCGGTGCTG 837
                 |||    |  | ||||   || |   |||||||| | |||  | |||||  || ||
Db       7454 ACCCTGGCCGTCGACGAGGGCTTCACGCACTTCGTCGAGGTCAGCGCCCACCCCGTCCTC 7513

Qy        838 ACCATGGGCCTCCAGGAGCTCGCCCCGGACCTGGGC 873
              ||||||  |||||  |||  || | | | ||| |||
Db       7514 ACCATGACCCTCCCCGAGACCGTCACCGGCCTCGGC 7549


RESULT 7
US-09-836-821-5
; Sequence 5, Application US/09836821
; Publication No. US20030087405A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/836,821
;  CURRENT FILING DATE:  2001-04-17
;  PRIOR APPLICATION NUMBER: 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 5
;   LENGTH: 36778
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-836-821-5

  Query Match             30.0%;  Score 284;  DB 9;  Length 36778;
  Best Local Similarity   60.5%;  Pred. No. 1.1e-56;
  Matches  530;  Conservative    0;  Mismatches  325;  Indels   21;  Gaps    3;

Qy          4 TTCGTTTTTCCCGGGCAGGGCCCGCAATGGCCGGGCATGGGAAGGGAACTTCTCGACGCT 63
              ||||| || ||||| |||||| |||| ||| | ||||||||     ||||| || ||| ||
Db       6689 TTCGTGTTCCCCGGTCAGGGCACGCAGTGGGCTGGCATGGGTGCCGAACTGCTGGACTCT 6748

Qy         64 TCCGACGTCTTCCGGGAGAGCGTCCGCGCCTGCGAAGCCGCGTTCGCGCCCTACGTCGAC 123
              ||||  || |||  || |  | |   ||  || || |||||| |  | || |||||||||
Db       6749 TCCGCGGTGTTCGCGGCGGCCATGGCCGAGTGTGAGGCCGCGCTGTCCCCGTACGTCGAC 6808

Qy        124 TGGTCGGTGGAGCAGGTGTTGCGGGACTCGCCGGACGCTCCCGGGCTGGACCGGGTGGAC 183
              ||||| |||||   ||  | ||| |  | || |  || |||  |||||| ||||| ||
Db       6809 TGGTCGCTGGAGGCCGTCGTACGGCAGGCCCCCGGTGCGCCCACGCTGGAGCGGGTCGAT 6868

Qy        184 GTCGTCCAGCCGACCCTGTTCGCCGTCATGATCTCCCTGGCCGCCCTCTGGCGCTCGCAA 243
              ||||| |||||      |||||||||||| |||| |||||    | | ||||      ||
Db       6869 GTCGTGCAGCCTGTGACGTTCGCCGTCATGGTCTCGCTGGCTCGCGTGTGGCAGCACCAC 6928

Qy        244 GGGGTCGAGCCGTGCGCGGTGCTGGGACACAGCCTGGGCGAGATCGCGGCAGCCCACGTC 303
              || ||   |||    |||||  | || |||   | ||||||||||| || ||  |||||
Db       6929 GGTGTGACGCCCCAGGCGGTCGTCGGCCACTCGCAGGGCGAGATCGCCGCCGCGTACGTC 6988

Qy        304 TCGGGAGGCCTGTCCCTGGCCGACGCCGCACGCGTGGTGACGCTTTGGAGCCAGGCACAG 363
              | |||| |||| ||||| |||||||| ||||| || || ||  | ||| || |
Db       6989 GCCGGAGCCCTGCCCCTGGACGACGCCGCCCGCGTCGTCACCCTGCGCAGCAAGTCCATC 7048

Qy        364 ACCAC---CCTTGCCGGGACCGGCGCGCTCGTCTCCGTCGCCGCCACGCCGGATGAGCTC 420
              || |   ||| |||| |  ||||   |  ||| |||    |   ||| |   ||
Db       7049 GCCGCCCACCTCGCCGGCAAGGGCGGCATGCTGTCCCTCGCGCTGAACGAGGACGCCGTC 7108

Qy        421 CTGCCCCGAATCGCTCCGTGGACCGAGGACAACCCGGCGCGGCTCGCCGTCGCAGCCGTC 480
              |||   ||| |                |   ||   |  ||||  ||||||| ||||||
Db       7109 CTGGAGCGACT--------------GAGTGACTTCGACGGGCTGTCCGTCGCCGCCGTC 7153

Qy        481 AACGGACCCCGGAGCACAGTCGTTTCCGGTGCCCGCGAGGCCGTCGCGGACCTGGTGGCC 540
              ||||| |||     ||| ||||| || |||| || ||     |||  || || |
Db       7154 AACGGGCCCACCGCCACTGTCGTGTCGGGTGACCCCGTACAGATCGAAGAGCTTGCTCAG 7213

Qy        541 GACCTCACCGCCGCGCAGGTGCGCACGCGCATGATCCCGGTGGACGTTCCCGCCCACTCC 600
              |    ||  || |     | ||| |||| || || || || |||    |  |||||  |
Db       7214 GCGTGCAAGGCGGACGGATTCCGCGCGCGGATCATTCCCGTCGACTACGCGTCCCACAGC 7273
```

```
Qy        601 CCCCTGATGTACGCCATCGAGGAACGGGTCGTCAGCGGCCTGCTGCCCATCACCCCACGC 660
              |  | | |   |   |||||||     | ||| |   | |||          ||| ||| |
Db       7274 CGGCAGGTCGAGATCATCGAGAGCGAGCTCGCCCAGGTCCTCGCCGGTCTCAGCCCGCAG 7333

Qy        661 CCCTCCCGCATCCCCTTCCACTCCTCGGTGACCGGCGGCCGCCTCGACACCCGCGAGCTA 720
              || | ||| | || |||  |||  || |    |||  | |  ||  |   | ||  ||
Db       7334 GCCCCGCGCGTGCCGTTCTTCTCGACGCTCGAAGGCACCTGGATCACCGAGCCCGTCCTC 7393

Qy        721 GACGCGGCGTACTGGTACCGCAACATGTCGAGCACGGTCCGGTTCGAGCCCGCCGCCCGG 780
              ||||   | ||||||||||||||| |      |   ||| | ||||  ||||||  |  |
Db       7394 GACGGCACCTACTGGTACCGCAACCTCCGTCACCGCGTCGGCTTCGCCCCCGCCATCGAG 7453

Qy        781 ---CTGCTTCTGCAGCAGGGGCCCAAGACGTTCGTCGAGATGAGCCCGCACCCGGTGCTG 837
                 |||    |  | ||||   || |   |||||||| | ||| | ||||| || ||
Db       7454 ACCCTGGCCGTCGACGAGGGCTTCACGCACTTCGTCGAGGTCAGCGCCCACCCCGTCCTC 7513

Qy        838 ACCATGGGCCTCCAGGAGCTCGCCCCGGACCTGGGC 873
              ||||||  |||||  |||  || | | | ||| |||
Db       7514 ACCATGACCCTCCCCGAGACCGTCACCGGCCTCGGC 7549


RESULT 8
US-09-861-289-5
; Sequence 5, Application US/09861289
; Patent No. US20020110897A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/861,289
;  CURRENT FILING DATE:  2001-05-18
;  PRIOR APPLICATION NUMBER: 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 5
;   LENGTH: 36778
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-861-289-5

  Query Match             30.0%;  Score 284;  DB 10;  Length 36778;
  Best Local Similarity   60.5%;  Pred. No. 1.1e-56;
  Matches  530;  Conservative    0;  Mismatches  325;  Indels   21;  Gaps    3;

Qy          4 TTCGTTTTTCCCGGGCAGGGCCCGCAATGGCCGGGCATGGGAAGGGAACTTCTCGACGCT 63
              ||||| || ||||| |||||| |||| ||| | ||||||||     ||||| |' ||| ||
Db       6689 TTCGTGTTCCCCGGTCAGGGCACGCAGTGGGCTGGCATGGGTGCCGAACTGCTGGACTCT 6748

Qy         64 TCCGACGTCTTCCGGGAGAGCGTCCGCGCCTGCGAAGCCGCGTTCGCGCCCTACGTCGAC 123
              ||||  || |||  || |  | |   ||  || || |||||| |  | || |||||||||
Db       6749 TCCGCGGTGTTCGCGGCGGCCATGGCCGAGTGTGAGGCCGCGCTGTCCCCGTACGTCGAC 6808

Qy        124 TGGTCGGTGGAGCAGGTGTTGCGGGACTCGCCGGACGCTCCCGGGCTGGACCGGGTGGAC 183
              |||||| |||||   ||  | ||| |  | || |  || |||  |||||| ||||| ||
Db       6809 TGGTCGCTGGAGGCCGTCGTACGGCAGGCCCCCGGTGCGCCCACGCTGGAGCGGGTCGAT 6868

Qy        184 GTCGTCCAGCCGACCCTGTTCGCCGTCATGATCTCCCTGGCCGCCCTCTGGCGCTCGCAA 243
              ||||| |||||      |||||||||||| |||| |||||   | | ||||      ||
Db       6869 GTCGTGCAGCCTGTGACGTTCGCCGTCATGGTCTCGCTGGCTCGCGTGTGGCAGCACCAC 6928

Qy        244 GGGGTCGAGCCGTGCGCGGTGCTGGGACACAGCCTGGGCGAGATCGCGGCAGCCCACGTC 303
              || ||   |||    |||||  | || |||   | ||||||||||||| || ||  |||||
Db       6929 GGTGTGACGCCCCAGGCGGTCGTCGGCCACTCGCAGGGCGAGATCGCCGCCGCGTACGTC 6988

Qy        304 TCGGGAGGCCTGTCCCTGGCCGACGCCGCACGCGTGGTGACGCTTTGGAGCCAGGCACAG 363
              | |||| |||| |||||| |||||||| ||||| || || ||  | ||| || |
Db       6989 GCCGGAGCCCTGCCCCTGGACGACGCCGCCCGCGTCGTCACCCTGCGCAGCAAGTCCATC 7048

Qy        364 ACCAC---CCTTGCCGGGACCGGCGCGCTCGTCTCCGTCGCCGCCACGCCGGATGAGCTC 420
              || |   ||| ||||| |  ||||   |  | ||| ||||    |   ||| |   ||
Db       7049 GCCGCCCACCTCGCCGGCAAGGGCGGCATGCTGTCCCTCGCGCTGAACGAGGACGCCGTC 7108

Qy        421 CTGCCCCGAATCGCTCCGTGGACCGAGGACAACCCGGCGCGGCTCGCCGTCGCAGCCGTC 480
              |||   ||| |                |   ||   |  ||||  ||||||| ||||||
Db       7109 CTGGAGCGACT---------------GAGTGACTTCGACGGGCTGTCCGTCGCCGCCGTC 7153

Qy        481 AACGGACCCCGGAGCACAGTCGTTTCCGGTGCCCGCGAGGCCGTCGCGGACCTGGTGGCC 540
              ||||| |||     ||| ||||| || |||| || ||       |||  || || |
Db       7154 AACGGGCCCACCGCCACTGTCGTGTCGGGTGACCCCGTACAGATCGAAGAGCTTGCTCAG 7213

Qy        541 GACCTCACCGCCGCGCAGGTGCGCACGCGCATGATCCCGGTGGACGTTCCCGCCCACTCC 600
              |    ||  || |      | ||| |||| || || || || |||    |  |||||  |
Db       7214 GCGTGCAAGGCGGACGGATTCCGCGCGCGGATCATTCCCGTCGACTACGCGTCCCACAGC 7273

Qy        601 CCCCTGATGTACGCCATCGAGGAACGGGTCGTCAGCGGCCTGCTGCCCATCACCCCACGC 660
              |  | | |   |   |||||||     | ||| |   | |||          ||| ||| |
Db       7274 CGGCAGGTCGAGATCATCGAGAGCGAGCTCGCCCAGGTCCTCGCCGGTCTCAGCCCGCAG 7333

Qy        661 CCCTCCCGCATCCCCTTCCACTCCTCGGTGACCGGCGGCCGCCTCGACACCCGCGAGCTA 720
              || | ||| | || |||  |||  || |    |||  | |  ||  |   | ||  ||
Db       7334 GCCCCGCGCGTGCCGTTCTTCTCGACGCTCGAAGGCACCTGGATCACCGAGCCCGTCCTC 7393

Qy        721 GACGCGGCGTACTGGTACCGCAACATGTCGAGCACGGTCCGGTTCGAGCCCGCCGCCCGG 780
              ||||   | ||||||||||||||| |      |   ||| | ||||  ||||||  |  |
Db       7394 GACGGCACCTACTGGTACCGCAACCTCCGTCACCGCGTCGGCTTCGCCCCCGCCATCGAG 7453

Qy        781 ---CTGCTTCTGCAGCAGGGGCCCAAGACGTTCGTCGAGATGAGCCCGCACCCGGTGCTG 837
                 |||    |  | ||||   || |   |||||||| | ||| | ||||| || ||
Db       7454 ACCCTGGCCGTCGACGAGGGCTTCACGCACTTCGTCGAGGTCAGCGCCCACCCCGTCCTC 7513

Qy        838 ACCATGGGCCTCCAGGAGCTCGCCCCGGACCTGGGC 873
              ||||||  |||||  |||  || | | | ||| |||
Db       7514 ACCATGACCCTCCCCGAGACCGTCACCGGCCTCGGC 7549


RESULT 9
US-09-988-384B-5
; Sequence 5, Application US/09988384B
; Publication No. US20030073824A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.536US1
;  CURRENT APPLICATION NUMBER: US/09/988,384B
;  CURRENT FILING DATE:  2001-11-19
;  PRIOR APPLICATION NUMBER: PCT/US99/14398
;  PRIOR FILING DATE: 1999-06-25
;  PRIOR APPLICATION NUMBER: US 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 53
; SEQ ID NO 5
;   LENGTH: 37948
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-988-384B-5

  Query Match             30.0%;  Score 284;  DB 9;  Length 37948;
  Best Local Similarity   60.5%;  Pred. No. 1.1e-56;
  Matches  530;  Conservative    0;  Mismatches  325;  Indels   21;  Gaps    3;

Qy          4 TTCGTTTTTCCCGGGCAGGGCCCGCAATGGCCGGGCATGGGAAGGGAACTTCTCGACGCT 63
              ||||| || ||||| |||||| |||| ||| | ||||||||     ||||| || ||| ||
Db       7859 TTCGTGTTCCCCGGTCAGGGCACGCAGTGGGCTGGCATGGGTGCCGAACTGCTGGACTCT 7918

Qy         64 TCCGACGTCTTCCGGGAGAGCGTCCGCGCCTGCGAAGCCGCGTTCGCGCCCTACGTCGAC 123
              ||||  || |||  || |  | |   ||  || || |||||| |  | || |||||||||
Db       7919 TCCGCGGTGTTCGCGGCGGCCATGGCCGAGTGTGAGGCCGCGCTGTCCCCGTACGTCGAC 7978

Qy        124 TGGTCGGTGGAGCAGGTGTTGCGGGACTCGCCGGACGCTCCCGGGCTGGACCGGGTGGAC 183
```

```
Db       7979 TGGTCGCTGGAGGCCGTCGTACGGCAGGCCCCCGGTGCGCCCACGCTGGAGCGGGTCGAT 8038

Qy        184 GTCGTCCAGCCGACCCTGTTCGCCGTCATGATCTCCCTGGCCGCCCTCTGGCGCTCGCAA 243

Db       8039 GTCGTGCAGCCTGTGACGTTCGCCGTCATGGTCTCGCTGGCTCGCGTGTGGCAGCACCAC 8098

Qy        244 GGGGTCGAGCCGTGCGCGGTGCTGGGACACAGCCTGGGCGAGATCGCGGCAGCCCACGTC 303

Db       8099 GGTGTGACGCCCCAGGCGGTCGTCGGCCACTCGCAGGGCGAGATCGCCGCCGCGTACGTC 8158

Qy        304 TCGGGAGGCCTGTCCCTGGCCGACGCCGCACGCGTGGTGACGCTTTGGAGCCAGGCACAG 363

Db       8159 GCCGGAGCCCTGCCCCTGGACGACGCCGCCCGCGTCGTCACCCTGCGCAGCAAGTCCATC 8218

Qy        364 ACCAC---CCTTGCCGGGACCGGCGCGCTCGTCTCCGTCGCCGCCACGCCGGATGAGCTC 420

Db       8219 GCCGCCCACCTCGCCGGCAAGGGCGGCATGCTGTCCCTCGCGCTGAACGAGGACGCCGTC 8278

Qy        421 CTGCCCCGAATCGCTCCGTGGACCGAGGACAACCCGGCGCGGCTCGCCGTCGCAGCCGTC 480

Db       8279 CTGGAGCGACT--------------GAGTGACTTCGACGGGCTGTCCGTCGCCGCCGTC 8323

Qy        481 AACGGACCCCGGAGCACAGTCGTTTCCGGTGCCCGCGAGGCCGTCGCGGACCTGGTGGCC 540

Db       8324 AACGGGCCCACCGCCACTGTCGTGTCGGGTGACCCCGTACAGATCGAAGAGCTTGCTCAG 8383

Qy        541 GACCTCACCGCCGCGCAGGTGCGCACGCGCATGATCCCGGTGGACGTTCCCGCCCACTCC 600

Db       8384 GCGTGCAAGGCGGACGGATTCCGCGCGCGGATCATTCCCGTCGACTACGCGTCCCACAGC 8443

Qy        601 CCCCTGATGTACGCCATCGAGGAACGGGTCGTCAGCGGCCTGCTGCCCATCACCCCACGC 660

Db       8444 CGGCAGGTCGAGATCATCGAGAGCGAGCTCGCCCAGGTCCTCGCCGGTCTCAGCCCGCAG 8503

Qy        661 CCCTCCCGCATCCCCTTCCACTCCTCGGTGACCGGCGGCCGCCTCGACACCCGCGAGCTA 720

Db       8504 GCCCCGCGCGTGCCGTTCTTCTCGACGCTCGAAGGCACCTGGATCACCGAGCCCGTCCTC 8563

Qy        721 GACGCGGCGTACTGGTACCGCAACATGTCGAGCACGGTCCGGTTCGAGCCCGCCGCCCGG 780

Db       8564 GACGGCACCTACTGGTACCGCAACCTCCGTCACCGCGTCGGCTTCGCCCCCGCCATCGAG 8623

Qy        781 ---CTGCTTCTGCAGCAGGGGCCCAAGACGTTCGTCGAGATGAGCCCGCACCCGGTGCTG 837

Db       8624 ACCCTGGCCGTCGACGAGGGCTTCACGCACTTCGTCGAGGTCAGCGCCCACCCCGTCCTC 8683

Qy        838 ACCATGGGCCTCCAGGAGCTCGCCCCGGACCTGGGC 873

Db       8684 ACCATGACCCTCCCCGAGACCGTCACCGGCCTCGGC 8719
```

```
RESULT 10
US-09-793-708-19
; Sequence 19, Application US/09793708
; Publication No. US20030104597A1
; GENERAL INFORMATION:
;  APPLICANT: ASHLEY, Gary
;  APPLICANT:  BETLACH, Melanie C.
;  APPLICANT:  BETLACH, Mary C.
;  APPLICANT:  McDANIEL, Robert
;  APPLICANT:  TANG, Li
;  TITLE OF INVENTION: RECOMBINANT NARBONOLIDE POLYKETIDE SYNTHASE
;  FILE REFERENCE: 300622002121
;  CURRENT APPLICATION NUMBER: US/09/793,708
;  CURRENT FILING DATE:  2001-02-22
;  PRIOR APPLICATION NUMBER: US 09/657,440
;  PRIOR FILING DATE: 2000-09-07
;  PRIOR APPLICATION NUMBER: US 09/320,878
;  PRIOR FILING DATE: 1999-05-27
;  PRIOR APPLICATION NUMBER: US 09/141,908
;  PRIOR FILING DATE: 1998-08-28
;  PRIOR APPLICATION NUMBER: US 09/073,538
;  PRIOR FILING DATE: 1998-05-06
;  PRIOR APPLICATION NUMBER: US 08/846,247
;  PRIOR FILING DATE: 1997-04-30
;  PRIOR APPLICATION NUMBER: US 60/134,990
;  PRIOR FILING DATE: 1999-05-20
;  NUMBER OF SEQ ID NOS: 38
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 19
;  LENGTH: 38506
;  TYPE: DNA
;  ORGANISM: Streptomyces venezuelae
US-09-793-708-19

  Query Match             30.0%;  Score 284;  DB 9;  Length 38506;
  Best Local Similarity   60.5%;  Pred. No. 1.1e-56;
  Matches  530;  Conservative    0;  Mismatches  325;  Indels   21;  Gaps    3;

Qy          4 TTCGTTTTTCCCGGGCAGGGCCCGCAATGGCCGGGCATGGGAAGGGAACTTCTCGACGCT 63

Db       4831 TTCGTGTTCCCCGGTCAGGGCACGCAGTGGGCTGGCATGGGTGCCGAACTGCTGGACTCT 4890

Qy         64 TCCGACGTCTTCCGGGAGAGCGTCCGCGCCTGCGAAGCCGCGTTCGCGCCCTACGTCGAC 123

Db       4891 TCCGCGGTGTTCGCGGCGGCCATGGCCGAGTGTGAGGCCGCGCTGTCCCCGTACGTCGAC 4950

Qy        124 TGGTCGGTGGAGCAGGTGTTGCGGGACTCGCCGGACGCTCCCGGGCTGGACCGGGTGGAC 183

Db       4951 TGGTCGCTGGAGGCCGTCGTACGGCAGGCCCCCGGTGCGCCCACGCTGGAGCGGGTCGAT 5010

Qy        184 GTCGTCCAGCCGACCCTGTTCGCCGTCATGATCTCCCTGGCCGCCCTCTGGCGCTCGCAA 243

Db       5011 GTCGTGCAGCCTGTGACGTTCGCCGTCATGGTCTCGCTGGCTCGCGTGTGGCAGCACCAC 5070

Qy        244 GGGGTCGAGCCGTGCGCGGTGCTGGGACACAGCCTGGGCGAGATCGCGGCAGCCCACGTC 303

Db       5071 GGTGTGACGCCCCAGGCGGTCGTCGGCCACTCGCAGGGCGAGATCGCCGCCGCGTACGTC 5130

Qy        304 TCGGGAGGCCTGTCCCTGGCCGACGCCGCACGCGTGGTGACGCTTTGGAGCCAGGCACAG 363

Db       5131 GCCGGAGCCCTGCCCCTGGACGACGCCGCCCGCGTCGTCACCCTGCGCAGCAAGTCCATC 5190

Qy        364 ACCAC---CCTTGCCGGGACCGGCGCGCTCGTCTCCGTCGCCGCCACGCCGGATGAGCTC 420

Db       5191 GCCGCCCACCTCGCCGGCAAGGGCGGCATGCTGTCCCTCGCGCTGAACGAGGACGCCGTC 5250

Qy        421 CTGCCCCGAATCGCTCCGTGGACCGAGGACAACCCGGCGCGGCTCGCCGTCGCAGCCGTC 480

Db       5251 CTGGAGCGACT--------------GAGTGACTTCGACGGGCTGTCCGTCGCCGCCGTC 5295

Qy        481 AACGGACCCCGGAGCACAGTCGTTTCCGGTGCCCGCGAGGCCGTCGCGGACCTGGTGGCC 540

Db       5296 AACGGGCCCACCGCCACTGTCGTGTCGGGTGACCCCGTACAGATCGAAGAGCTTGCTCAG 5355

Qy        541 GACCTCACCGCCGCGCAGGTGCGCACGCGCATGATCCCGGTGGACGTTCCCGCCCACTCC 600

Db       5356 GCGTGCAAGGCGGACGGATTCCGCGCGCGGATCATTCCCGTCGACTACGCGTCCCACAGC 5415

Qy        601 CCCCTGATGTACGCCATCGAGGAACGGGTCGTCAGCGGCCTGCTGCCCATCACCCCACGC 660

Db       5416 CGGCAGGTCGAGATCATCGAGAGCGAGCTCGCCCAGGTCCTCGCCGGTCTCAGCCCGCAG 5475

Qy        661 CCCTCCCGCATCCCCTTCCACTCCTCGGTGACCGGCGGCCGCCTCGACACCCGCGAGCTA 720

Db       5476 GCCCCGCGCGTGCCGTTCTTCTCGACGCTCGAAGGCACCTGGATCACCGAGCCCGTCCTC 5535

Qy        721 GACGCGGCGTACTGGTACCGCAACATGTCGAGCACGGTCCGGTTCGAGCCCGCCGCCCGG 780

Db       5536 GACGGCACCTACTGGTACCGCAACCTCCGTCACCGCGTCGGCTTCGCCCCCGCCATCGAG 5595

Qy        781 ---CTGCTTCTGCAGCAGGGGCCCAAGACGTTCGTCGAGATGAGCCCGCACCCGGTGCTG 837

Db       5596 ACCCTGGCCGTCGACGAGGGCTTCACGCACTTCGTCGAGGTCAGCGCCCACCCCGTCCTC 5655
```

```
Qy        838 ACCATGGGCCTCCAGGAGCTCGCCCCGGACCTGGGC 873
              |||||| ||||| ||| || | | | ||| |||
Db       5656 ACCATGACCCTCCCCGAGACCGTCACCGGCCTCGGC 5691


RESULT 11
US-09-735-056-30
; Sequence 30, Application US/09735056
; Publication No. US20030013662A1
;  GENERAL INFORMATION:
;    APPLICANT:  Katz, Leonard
;    APPLICANT:  Stassi, Diane L.
;    APPLICANT:  Summers Jr., Richard G.
;    APPLICANT:  Ruan, Xiaoan
;    APPLICANT:  Pereda-Lopez, Ana
;    APPLICANT:  Kakavas, Stephan J.
;    TITLE OF INVENTION:  NOVEL POLYKETIDE DERIVATIVES
;    TITLE OF INVENTION:  AND RECOMBINANT METHODS FOR MAKING SAME
;    NUMBER OF SEQUENCES:  34
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Abbott Laboratories
;      STREET:  100 Abbott Park Rd.
;      CITY:  Abbott Park
;      STATE:  Illinois
;      COUNTRY:  USA
;      ZIP:  60064-3500
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/735,056
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/858,003
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Dianne Casuto
;      REGISTRATION NUMBER:  P-40,943
;      REFERENCE/DOCKET NUMBER:  4952.US.P2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (847)-938-3137
;      TELEFAX:  (847)-938-2623
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  30:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1035 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
US-09-735-056-30

  Query Match           29.5%;  Score 279.6;  DB 9;  Length 1035;
  Best Local Similarity   58.1%;  Pred. No. 2e-55;
  Matches  539;  Conservative    0;  Mismatches  374;  Indels   15;  Gaps    2;

Qy          1 GTCTTCGTTTTTCCCGGGCAGGGCCCGCAATGGCCGGGCATGGGAAGGGAACTTCTCGAC 60
              | |||||| |||||||||||||||| | || ||| | ||  |||||  |    || ||||
Db         10 GCCTTCGTCTTTCCCGGGCAGGGCGCCCAGTGGGCCGGACTGGGAGCGCGGCTCCTCGCG 69

Qy         61 GCTTCCGACGTCTTCCGGGAGAGCGTCCGCGCCTGCGAAGCCGCGTTCGCGCCCTACGTC 120
              |  |||  |||||||||  |  || | |   || ||||     ||| | | |||  || ||
Db         70 GACTCCCCCGTCTTCCGCGCCAGGGCCGAGGCATGCGCGCGGGCGCTGGAGCCTCACCTC 129

Qy        121 GACTGGTCGGTGGAGCAGGTGTTGCGGGACTCGCCGGACGCTCCCGGGCTGGACCGGGTG 180
              ||||||||||      | ||| ||   | | | |||| | | ||      | |||||||
Db        130 GACTGGTCGGTCCTCGACGTGCTGGCCGGCGCCCCGGGCACCCCTCCCATCGACCGGGCC 189

Qy        181 GACGTCGTCCAGCCGACCCTGTTCGCCGTCATGATCTCCCTGGCCGCCCTCTGGCGCTCG 240
              |||||  || ||||||   |||||| ||   ||| |||| ||||||||||||||||    |
Db        190 GACGTGGTGCAGCCGGTGCTGTTCACCACGATGGTCTCGCTGGCCGCCCTCTGGGAGGCC 249

Qy        241 CAAGGGGTCGAGCCGTGCGCGGTGCTGGGACACAGCCTGGGCGAGATCGCGGCAGCCCAC 300
              || |||||   ||||  |||||| |||| |||  || ||||||| | || || |||  |
Db        250 CACGGGGTGCGGCCGGCCGCGGTCGTGGGCCACTCCCAGGGCGAGGTGGCCGCGGCCTGC 309

Qy        301 GTCTCGGGAGGCCTGTCCCTGGCCGACGCCGCACGCGTGGTGACGCTTTGGAGCCAGGCA 360
              ||  | || | ||||| |||| |||||| || |  ||| |  |     | |||  |
Db        310 GTGGCCGGTGCCCTGTCGCTGGACGACGCTGCCCTGGTGATCGCCGGACGCAGCAGGCTG 369

Qy        361 CAGACCACCCTTGCCGGGACCGGCGCGCTCGTCTCCGTCGCCGCCACGCCGGATGAGCTC 420
                |       || ||||||| |||| | |  || | ||    ||  || | ||   | ||
Db        370 TGGGGGCGGCTGGCCGGGAACGGCGGGATGCTCGCGGTGATGGCTCCGGCCGAGCGGATC 429

Qy        421 CTGCCCCGAATCGCTCCGTGGACCGAGGACAACCCGGCGCGGCTCGCCGTCGCAGCCGTC 480
              |     |  |||  || |||            | |  |||| |  | || || || |||
Db        430 CGTGAGCTGCTCGAACCATGG-----------CGGCAGCGGATTTCGGTGGCGGCGGTC 477

Qy        481 AACGGACCCCGGAGCACAGTCGTTTCCGGTGCCCGCGAGGCCGTCGCGGACCTGGTGGCC 540
              || || |||          ||| ||||||| |      ||  | | |||  | |  ||
Db        478 AATGGCCCCGCCTCGGTCACCGTCTCCGGTGACGCGCTCGCGCTGGAGGAGTTCGGCGCG 537

Qy        541 GACCTCACCGCCGCGCAGG---TGCGCACGCGCATGATCCCGGTGGACGTTCCCGCCCAC 597
                 ||| |||||| |  ||   |||||  ||   ||       || ||| |  ||| ||||
Db        538 CGGCTCTCCGCCGAGGGGGTGCTGCGCTGGCCGCTGCCGGGCGTCGACTTCGCCGGCCAC 597

Qy        598 TCCCCCCTGATGTACGCCATCGAGGAACGGGTCGTCAGCGGCCTGCTGCCCATCACCCCA 657
              || || | | || | |   ||   |    | || |   |   ||    | |    ||
Db        598 TCGCCGCAGGTGGAGGAGTTCCGCGCTGAGCTCCTGGACCTGCTCTCCGGCGTACGGCCG 657

Qy        658 CGCCCCTCCCGCATCCCCTTCCACTCCTCGGTGACCGGCGGCCGCCTCGACACCCGCGAG 717
                 || || || || || |||  |||| | ||||| |  || |    || |  |  | ||
Db        658 GCTCCTTCGCGGATACCTTTCTTCTCCACCGTGACGGCGGGTCCTTGCGGCGGCGACCAG 717

Qy        718 CTAGACGCGGCGTACTGGTACCGCAACATGTCGAGCACGGTCCGGTTCGAGCCCGCCGCC 777
              || |||| |||||||||||||||||||| |      | ||   ||||||  || | | |
Db        718 CTGGACGGGGCGTACTGGTACCGCAACACGCGCGAACCCGTGGAGTTCGACGCCACGGTC 777

Qy        778 CGGCTGCTTCTGCAGCAGGGGCCCAAGACGTTCGTCGAGATGAGCCCGCACCCGGTGCTG 837
              |||  ||| ||||     ||| |   | |||||| ||||| |  | ||||| ||| ||||
Db        778 CGGGCGCTGCTGCGTGCGGGCCATCACACGTTCATCGAGGTCGGTCCGCATCCGCTGCTC 837

Qy        838 ACCATGGGCCTCCAGGAGCTCGCCCCGGACCTGGGCGACACCACCGGCACCGCCGACACC 897
              | |   || | ||| |||| ||||| ||| |   |  ||    ||    | ||
Db        838 AACGCCGCGATCGACGAGATCGCAGCGGACGAGGGGGTAGCGGCCACGGCCCTGCATACG 897

Qy        898 GTGATCATGGGCACGCTGCGCCGCGGCC 925
              |      |||| |     ||| | ||
Db        898 CTCCAGCGGGGCGCTGGCGGCCTTGACC 925


RESULT 12
US-09-860-846-32
; Sequence 32, Application US/09860846
; Patent No. US20020164742A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/860,846
;  CURRENT FILING DATE:  2001-05-18
;  PRIOR APPLICATION NUMBER: 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 32
;    LENGTH: 11220
;    TYPE: DNA
;    ORGANISM: Streptomyces venezuelae
US-09-860-846-32
```

```
  Query Match           28.2%;  Score 267.2;  DB 9;  Length 11220;
  Best Local Similarity  59.0%;  Pred. No. 9.4e-53;
  Matches  524;  Conservative    0;  Mismatches  343;  Indels   21;  Gaps    3;

Qy          4 TTCGTTTTTCCCGGGCAGGGCCCGCAATGGCCGGGCATGGGAAGGGAACTTCTCGACGCT 63
              ||||| || ||||| |||||| |||| ||| | || |||||    ||||| |||||||
Db       1714 TTCGTGTTCCCCGGTCAGGGCACGCAGTGGGCCGGGATGGGCGCCGAACTCCTCGACGTG 1773

Qy         64 TCCGACGTCTTCCGGGAGAGCGTCCGCGCCTGCGAAGCCGCGTTCGCGCCCTACGTCGAC 123
              ||  | |  ||| || | |   || |||||   |||| || | | ||| ||||||
Db       1774 TCGAAGGAGTTCGCGGCGGCCATGGCCGAGTGCGAGAGCGCGCTCTCCCGCTATGTCGAC 1833

Qy        124 TGGTCGGTGGAGCAGGTGTTGCGGGACTCGCCGGACGCTCCCGGGCTGGACCGGGTGGAC 183
              |||||| |||||   ||  | ||| |  |||||| ||| |||  |||||| ||||| |||
Db       1834 TGGTCGCTGGAGGCCGTCGTCCGGCAGGCGCCGGGCGCGCCCACGCTGGAGCGGGTCGAC 1893

Qy        184 GTCGTCCAGCCGACCCTGTTCGCCGTCATGATCTCCCTGGCCGCCCTCTGGCGCTCGCAA 243
              ||||||||||       ||||| |||||| | || |||||     ||||||      ||
Db       1894 GTCGTCCAGCCCGTGACCTTCGCTGTCATGGTTTCGCTGGCGAAGGTCTGGCAGCACCAC 1953

Qy        244 GGGGTCGAGCCGTGCGCGGTGCTGGGACACAGCCTGGGCGAGATCGCGGCAGCCCACGTC 303
              || ||   ||||   || ||  | || |||    | ||||||||||| || ||  |||||
Db       1954 GGCGTGACGCCGCAGGCCGTCGTCGGCCACTCGCAGGGCGAGATCGCCGCCGCGTACGTC 2013

Qy        304 TCGGGAGGCCTGTCCCTGGCCGACGCCGCACGCGTGGTGACGCTTTGGAGCCAGGCACAG 363
              | || | |||  |||| | |||||||| ||||| || || ||  | ||| || |
Db       2014 GCCGGTGCCCTCACCCTCGACGACGCCGCCCGCGTCGTCACCCTGCGCAGCAAGTCCATC 2073

Qy        364 ACCAC---CCTTGCCGGGACCGGCGCGCTCGTCTCCGTCGCCGCCACGCCGGATGAGCTC 420
              || |   ||| ||||| |  ||||   | |||||  |||||  ||     ||| |    |
Db       2074 GCCGCCCACCTCGCCGGCAAGGGCGGCATGATCTCCCTCGCCCTCAGCGAGGAAGCCACC 2133

Qy        421 CTGCCCCGAATCGCTCCGTGGACCGAGGACAACCCGGCGCGGCTCGCCGTCGCAGCCGTC 480
              | ||  || |||          || ||||       | ||  |  |||| ||||||
Db       2134 CGGCAGCGCATC--------------GAGAACCTCCACGGACTGTCGATCGCCGCCGTC 2178

Qy        481 AACGGACCCCGGAGCACAGTCGTTTCCGGTGCCCGCGAGGCCGTCGCGGACCTGGTGGCC 540
              ||||| |||     ||| || ||||| || | || |       ||   || || |
Db       2179 AACGGCCCCACCGCCACCGTGGTTTCGGGCGACCCCACCCAGATCCAAGAGCTCGCTCAG 2238

Qy        541 GACCTCACCGCCGCGCAGGTGCGCACGCGCATGATCCCGGTGGACGTTCCCGCCCACTCC 600
              |      ||||    ||| ||| | || || ||||| || |||    || |||||  |
Db       2239 GCGTGTGAGGCCGACGGGGTCCGCGCACGGATCATCCCCGTCGACTACGCCTCCCACAGC 2298

Qy        601 CCCCTGATGTACGCCATCGAGGAACGGGTCGTCAGCGGCCTGCTGCCCATCACCCCACGC 660
              |||  |  | ||||||||          ||| |   | |||          ||| ||| ||
Db       2299 GCCCACGTCGAGACCATCGAGAGCGAACTCGCCGAGGTCCTCGCCGGGCTCAGCCCGCGG 2358

Qy        661 CCCTCCCGCATCCCCTTCCACTCCTCGGTGACCGGCGGCCGCCTCGACACCCGCGAGCTA 720
              | |   | || |||  ||| |  |    |||| | |   ||  |   |  | |||
Db       2359 ACACCTGAGGTGCCGTTCTTCTCGACACTCGAAGGCGCCTGGATCACCGAGCCGGTGCTC 2418

Qy        721 GACGCGGCGTACTGGTACCGCAACATGTCGAGCACGGTCCGGTTCGAGCCCGCCGCCCGG 780
              ||||   | ||||||||||||||| |       |||  | ||||  |||||||  |  |
Db       2419 GACGGCACCTACTGGTACCGCAACCTCCGCCACCGCGTCGGCTTCGCCCCCGCCGTCGAG 2478

Qy        781 ---CTGCTTCTGCAGCAGGGGCCCAAGACGTTCGTCGAGATGAGCCCGCACCCGGTGCTG 837
                 ||      |  ||    ||      ||| ||||| | ||| | ||||| || ||
Db       2479 ACCCTCGCCACCGACGAAGGCTTCACCCACTTCATCGAGGTCAGCGCCCACCCGTCCTC 2538

Qy        838 ACCATGGGCCTCCAGGAGCTCGCCCCGGACCTGGGCGACACCACCGGC 885
              ||||||  |||||  |||  || | | ||| |||  |  |  | ||
Db       2539 ACCATGACCCTCCCCGAGACCGTCACCGGCCTCGGCACCCTCCGCCGC 2586


RESULT 13
US-09-988-384B-32
; Sequence 32, Application US/09988384B
; Publication No. US20030073824A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.536US1
;  CURRENT APPLICATION NUMBER: US/09/988,384B
;  CURRENT FILING DATE:  2001-11-19
;  PRIOR APPLICATION NUMBER: PCT/US99/14398
;  PRIOR FILING DATE: 1999-06-25
;  PRIOR APPLICATION NUMBER: US 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 53
; SEQ ID NO 32
;   LENGTH: 11220
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-988-384B-32

  Query Match           28.2%;  Score 267.2;  DB 9;  Length 11220;
  Best Local Similarity  59.0%;  Pred. No. 9.4e-53;
  Matches  524;  Conservative    0;  Mismatches  343;  Indels   21;  Gaps    3;

Qy          4 TTCGTTTTTCCCGGGCAGGGCCCGCAATGGCCGGGCATGGGAAGGGAACTTCTCGACGCT 63
              ||||| || ||||| |||||| |||| ||| | || |||||    ||||| |||||||
Db       1714 TTCGTGTTCCCCGGTCAGGGCACGCAGTGGGCCGGGATGGGCGCCGAACTCCTCGACGTG 1773

Qy         64 TCCGACGTCTTCCGGGAGAGCGTCCGCGCCTGCGAAGCCGCGTTCGCGCCCTACGTCGAC 123
              ||  | |  ||| || | |   || |||||   |||| || | | ||| ||||||
Db       1774 TCGAAGGAGTTCGCGGCGGCCATGGCCGAGTGCGAGAGCGCGCTCTCCCGCTATGTCGAC 1833

Qy        124 TGGTCGGTGGAGCAGGTGTTGCGGGACTCGCCGGACGCTCCCGGGCTGGACCGGGTGGAC 183
              |||||| |||||   ||  | ||| |  |||||| ||| |||  |||||| ||||| |||
Db       1834 TGGTCGCTGGAGGCCGTCGTCCGGCAGGCGCCGGGCGCGCCCACGCTGGAGCGGGTCGAC 1893

Qy        184 GTCGTCCAGCCGACCCTGTTCGCCGTCATGATCTCCCTGGCCGCCCTCTGGCGCTCGCAA 243
              ||||||||||       ||||| |||||| | || |||||     ||||||      ||
Db       1894 GTCGTCCAGCCCGTGACCTTCGCTGTCATGGTTTCGCTGGCGAAGGTCTGGCAGCACCAC 1953

Qy        244 GGGGTCGAGCCGTGCGCGGTGCTGGGACACAGCCTGGGCGAGATCGCGGCAGCCCACGTC 303
              || ||   ||||   || ||  | || |||    | ||||||||||| || ||  |||||
Db       1954 GGCGTGACGCCGCAGGCCGTCGTCGGCCACTCGCAGGGCGAGATCGCCGCCGCGTACGTC 2013

Qy        304 TCGGGAGGCCTGTCCCTGGCCGACGCCGCACGCGTGGTGACGCTTTGGAGCCAGGCACAG 363
              | || | |||  |||| | |||||||| ||||| || || ||  | ||| || |
Db       2014 GCCGGTGCCCTCACCCTCGACGACGCCGCCCGCGTCGTCACCCTGCGCAGCAAGTCCATC 2073

Qy        364 ACCAC---CCTTGCCGGGACCGGCGCGCTCGTCTCCGTCGCCGCCACGCCGGATGAGCTC 420
              || |   ||| ||||| |  ||||   | |||||  |||||  ||     ||| |    |
Db       2074 GCCGCCCACCTCGCCGGCAAGGGCGGCATGATCTCCCTCGCCCTCAGCGAGGAAGCCACC 2133

Qy        421 CTGCCCCGAATCGCTCCGTGGACCGAGGACAACCCGGCGCGGCTCGCCGTCGCAGCCGTC 480
              | ||  || |||          || ||||       | ||  |  |||| ||||||
Db       2134 CGGCAGCGCATC--------------GAGAACCTCCACGGACTGTCGATCGCCGCCGTC 2178

Qy        481 AACGGACCCCGGAGCACAGTCGTTTCCGGTGCCCGCGAGGCCGTCGCGGACCTGGTGGCC 540
              ||||| |||     ||| || ||||| || | || |       ||   || || |
Db       2179 AACGGCCCCACCGCCACCGTGGTTTCGGGCGACCCCACCCAGATCCAAGAGCTCGCTCAG 2238

Qy        541 GACCTCACCGCCGCGCAGGTGCGCACGCGCATGATCCCGGTGGACGTTCCCGCCCACTCC 600
              |      ||||    ||| ||| | || || ||||| || |||    || |||||  |
Db       2239 GCGTGTGAGGCCGACGGGGTCCGCGCACGGATCATCCCCGTCGACTACGCCTCCCACAGC 2298

Qy        601 CCCCTGATGTACGCCATCGAGGAACGGGTCGTCAGCGGCCTGCTGCCCATCACCCCACGC 660
              |||  |  | ||||||||          ||| |   | |||          ||| ||| ||
Db       2299 GCCCACGTCGAGACCATCGAGAGCGAACTCGCCGAGGTCCTCGCCGGGCTCAGCCCGCGG 2358

Qy        661 CCCTCCCGCATCCCCTTCCACTCCTCGGTGACCGGCGGCCGCCTCGACACCCGCGAGCTA 720
              | |   | || |||  ||| |  |    |||| | |   ||  |   |  | |||
Db       2359 ACACCTGAGGTGCCGTTCTTCTCGACACTCGAAGGCGCCTGGATCACCGAGCCGGTGCTC 2418

Qy        721 GACGCGGCGTACTGGTACCGCAACATGTCGAGCACGGTCCGGTTCGAGCCCGCCGCCCGG 780
              ||||   | ||||||||||||||| |       |||  | ||||  |||||||  |  |
Db       2419 GACGGCACCTACTGGTACCGCAACCTCCGCCACCGCGTCGGCTTCGCCCCCGCCGTCGAG 2478

Qy        781 ---CTGCTTCTGCAGCAGGGGCCCAAGACGTTCGTCGAGATGAGCCCGCACCCGGTGCTG 837
```

```
                ||          |  | ||   ||      ||| ||||| | ||| | ||||| || ||
Db       2479 ACCCTCGCCACCGACGAAGGCTTCACCCACTTCATCGAGGTCAGCGCCCACCCCGTCCTC 2538

Qy        838 ACCATGGGCCTCCAGGAGCTCGCCCCGGACCTGGGCGACACCACCGGC 885
              ||||||  |||||  |||  || | | | ||| |||  |  |  | ||
Db       2539 ACCATGACCCTCCCCGAGACCGTCACCGGCCTCGGCACCCTCCGCCGC 2586


RESULT 14
US-09-836-821-32
; Sequence 32, Application US/09836821
; Publication No. US20030087405A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/836,821
;  CURRENT FILING DATE:  2001-04-17
;  PRIOR APPLICATION NUMBER: 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 32
;   LENGTH: 11220
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-836-821-32

  Query Match             28.2%;  Score 267.2;  DB 9;  Length 11220;
  Best Local Similarity   59.0%;  Pred. No. 9.4e-53;
  Matches  524;  Conservative    0;  Mismatches  343;  Indels   21;  Gaps    3;

Qy          4 TTCGTTTTTCCCGGGCAGGGCCCGCAATGGCCGGGCATGGGAAGGGAACTTCTCGACGCT 63
              ||||| || ||||| |||||| |||| ||| | || |||||    ||||| |||||||
Db       1714 TTCGTGTTCCCCGGTCAGGGCACGCAGTGGGCCGGGATGGGCGCCGAACTCCTCGACGTG 1773

Qy         64 TCCGACGTCTTCCGGGAGAGCGTCCGCGCCTGCGAAGCCGCGTTCGCGCCCTACGTCGAC 123
              ||  | |  ||| || |  | |   ||  |||||   |||| || | | ||| ||||||
Db       1774 TCGAAGGAGTTCGCGGCGGCCATGGCCGAGTGCGAGAGCGCGCTCTCCCGCTATGTCGAC 1833

Qy        124 TGGTCGGTGGAGCAGGTGTTGCGGGACTCGCCGGACGCTCCCGGGCTGGACCGGGTGGAC 183
              |||||| |||||   ||  | ||| |  |||||| ||| |||  |||||| ||||| |||
Db       1834 TGGTCGCTGGAGGCCGTCGTCCGGCAGGCGCCGGGCGCGCCCACGCTGGAGCGGGTCGAC 1893

Qy        184 GTCGTCCAGCCGACCCTGTTCGCCGTCATGATCTCCCTGGCCGCCCTCTGGCGCTCGCAA 243
              ||||||||||       ||||| |||||| | || |||||     ||||||      ||
Db       1894 GTCGTCCAGCCCGTGACCTTCGCTGTCATGGTTTCGCTGGCGAAGGTCTGGCAGCACCAC 1953

Qy        244 GGGGTCGAGCCGTGCGCGGTGCTGGGACACAGCCTGGGCGAGATCGCGGCAGCCCACGTC 303
              || ||   ||||   || ||  | || |||   | ||||||||||| || ||  |||||
Db       1954 GGCGTGACGCCGCAGGCCGTCGTCGGCCACTCGCAGGGCGAGATCGCCGCCGCGTACGTC 2013

Qy        304 TCGGGAGGCCTGTCCCTGGCCGACGCCGCACGCGTGGTGACGCTTTGGAGCCAGGCACAG 363
              | || | |||  |||| | |||||||| |||||  || ||  ||  | ||| || |
Db       2014 GCCGGTGCCCTCACCCTCGACGACGCCGCCCGCGTCGTCACCCTGCGCAGCAAGTCCATC 2073

Qy        364 ACCAC---CCTTGCCGGGACCGGCGCGCTCGTCTCCGTCGCCGCCACGCCGGATGAGCTC 420
              || |   ||| |||||  |  ||||    | |||||  |||||   ||    ||| |    |
Db       2074 GCCGCCCACCTCGCCGGCAAGGGCGGCATGATCTCCCTCGCCCTCAGCGAGGAAGCCACC 2133

Qy        421 CTGCCCCGAATCGCTCCGTGGACCGAGGACAACCCGGCGCGGCTCGCCGTCGCAGCCGTC 480
              | ||  ||  |||                ||  ||||      | ||  |  ||||  ||||||
Db       2134 CGGCAGCGCATC---------------GAGAACCTCCACGGACTGTCGATCGCCGCCGTC 2178

Qy        481 AACGGACCCCGGAGCACAGTCGTTTCCGGTGCCCGCGAGGCCGTCGCGGACCTGGTGGCC 540
              |||||  |||     |||  || |||||  || |  || |        ||   ||  || |
Db       2179 AACGGCCCCACCGCCACCGTGGTTTCGGGCGACCCCACCCAGATCCAAGAGCTCGCTCAG 2238

Qy        541 GACCTCACCGCCGCGCAGGTGCGCACGCGCATGATCCCGGTGGACGTTCCCGCCCACTCC 600
              .  |        ||||    ||| ||| | || || ||||| || |||    || |||||  |
```

```
Db       2239 GCGTGTGAGGCCGACGGGGTCCGCGCACGGATCATCCCCGTCGACTACGCCTCCCACAGC 2298

Qy        601 CCCCTGATGTACGCCATCGAGGAACGGGTCGTCAGCGGCCTGCTGCCCATCACCCCACGC 660
              |||    |  |  ||||||||        ||| |    | |||         ||| ||| ||
Db       2299 GCCCACGTCGAGACCATCGAGAGCGAACTCGCCGAGGTCCTCGCCGGGCTCAGCCCGCGG 2358

Qy        661 CCCTCCCGCATCCCCTTCCACTCCTCGGTGACCGGCGGCCGCCTCGACACCCGCGAGCTA 720
              |  |      | || |||  |||  |  |      |||| | |  ||  |    |  | |||
Db       2359 ACACCTGAGGTGCCGTTCTTCTCGACACTCGAAGGCGCCTGGATCACCGAGCCGGTGCTC 2418

Qy        721 GACGCGGCGTACTGGTACCGCAACATGTCGAGCACGGTCCGGTTCGAGCCCGCCGCCCGG 780
              ||||   | ||||||||||||||| |      |    ||| | ||||  |||||| |  |
Db       2419 GACGGCACCTACTGGTACCGCAACCTCCGCCACCGCGTCGGCTTCGCCCCCGCCGTCGAG 2478

Qy        781 ---CTGCTTCTGCAGCAGGGGCCCAAGACGTTCGTCGAGATGAGCCCGCACCCGGTGCTG 837
                 ||         |  | ||   ||      ||| ||||| | ||| | ||||| || ||
Db       2479 ACCCTCGCCACCGACGAAGGCTTCACCCACTTCATCGAGGTCAGCGCCCACCCCGTCCTC 2538

Qy        838 ACCATGGGCCTCCAGGAGCTCGCCCCGGACCTGGGCGACACCACCGGC 885
              ||||||  |||||  |||  || | | | ||| |||  |  |  | ||
Db       2539 ACCATGACCCTCCCCGAGACCGTCACCGGCCTCGGCACCCTCCGCCGC 2586


RESULT 15
US-09-861-289-32
; Sequence 32, Application US/09861289
; Patent No. US20020110897A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/861,289
;  CURRENT FILING DATE:  2001-05-18
;  PRIOR APPLICATION NUMBER: 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 32
;   LENGTH: 11220
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-861-289-32

  Query Match             28.2%;  Score 267.2;  DB 10;  Length 11220;
  Best Local Similarity   59.0%;  Pred. No. 9.4e-53;
  Matches  524;  Conservative    0;  Mismatches  343;  Indels   21;  Gaps    3;

Qy          4 TTCGTTTTTCCCGGGCAGGGCCCGCAATGGCCGGGCATGGGAAGGGAACTTCTCGACGCT 63
              ||||| || ||||| |||||| |||| ||| | || |||||    ||||| |||||||
Db       1714 TTCGTGTTCCCCGGTCAGGGCACGCAGTGGGCCGGGATGGGCGCCGAACTCCTCGACGTG 1773

Qy         64 TCCGACGTCTTCCGGGAGAGCGTCCGCGCCTGCGAAGCCGCGTTCGCGCCCTACGTCGAC 123
              ||  | |  ||| || |  | |   ||  |||||   |||| || | | ||| ||||||
Db       1774 TCGAAGGAGTTCGCGGCGGCCATGGCCGAGTGCGAGAGCGCGCTCTCCCGCTATGTCGAC 1833

Qy        124 TGGTCGGTGGAGCAGGTGTTGCGGGACTCGCCGGACGCTCCCGGGCTGGACCGGGTGGAC 183
              |||||| |||||   ||  | ||| |  |||||| ||| |||  |||||| ||||| |||
Db       1834 TGGTCGCTGGAGGCCGTCGTCCGGCAGGCGCCGGGCGCGCCCACGCTGGAGCGGGTCGAC 1893

Qy        184 GTCGTCCAGCCGACCCTGTTCGCCGTCATGATCTCCCTGGCCGCCCTCTGGCGCTCGCAA 243
              ||||||||||       ||||| |||||| | || |||||     ||||||      ||
Db       1894 GTCGTCCAGCCCGTGACCTTCGCTGTCATGGTTTCGCTGGCGAAGGTCTGGCAGCACCAC 1953

Qy        244 GGGGTCGAGCCGTGCGCGGTGCTGGGACACAGCCTGGGCGAGATCGCGGCAGCCCACGTC 303
              || ||   ||||   || ||  | || |||   | ||||||||||| || ||  |||||
Db       1954 GGCGTGACGCCGCAGGCCGTCGTCGGCCACTCGCAGGGCGAGATCGCCGCCGCGTACGTC 2013

Qy        304 TCGGGAGGCCTGTCCCTGGCCGACGCCGCACGCGTGGTGACGCTTTGGAGCCAGGCACAG 363
              | || | |||  |||| | |||||||| |||||  || ||  ||  | ||| || |
Db       2014 GCCGGTGCCCTCACCCTCGACGACGCCGCCCGCGTCGTCACCCTGCGCAGCAAGTCCATC 2073
```

```
Qy        364 ACCAC---CCTTGCCGGGACCGGCGCGCTCGTCTCCGTCGCCGCCACGCCGGATGAGCTC 420
              || |   ||| ||||| |  ||||    |  ||||| |||||  ||    ||| |     |
Db       2074 GCCGCCCACCTCGCCGGCAAGGGCGGCATGATCTCCCTCGCCCTCAGCGAGGAAGCCACC 2133

Qy        421 CTGCCCCGAATCGCTCCGTGGACCGAGGACAACCCGGCGCGGCTCGCCGTCGCAGCCGTC 480
              | ||  || |||                || ||||       | ||   |  |||| ||||||
Db       2134 CGGCAGCGCATC--------------GAGAACCTCCACGGACTGTCGATCGCCGCCGTC 2178

Qy        481 AACGGACCCCGGAGCACAGTCGTTTCCGGTGCCCGCGAGGCCGTCGCGGACCTGGTGGCC 540
              ||||| |||     ||| || ||||| || | || |        ||   || || |
Db       2179 AACGGCCCCACCGCCACCGTGGTTTCGGGCGACCCCACCCAGATCCAAGAGCTCGCTCAG 2238

Qy        541 GACCTCACCGCCGCGCAGGTGCGCACGCGCATGATCCCGGTGGACGTTCCCGCCCACTCC 600
              |         ||||    ||| ||| | || || ||||| || |||    || |||||  |
Db       2239 GCGTGTGAGGCCGACGGGGTCCGCGCACGGATCATCCCCGTCGACTACGCCTCCCACAGC 2298

Qy        601 CCCCTGATGTACGCCATCGAGGAACGGGTCGTCAGCGGCCTGCTGCCCATCACCCCACGC 660
              |||   |  |  ||||||||       ||| |   | |||        ||| ||| ||
Db       2299 GCCCACGTCGAGACCATCGAGAGCGAACTCGCCGAGGTCCTCGCCGGGCTCAGCCCGCGG 2358

Qy        661 CCCTCCCGCATCCCCTTCCACTCCTCGGTGACCGGCGGCCGCCTCGACACCCGCGAGCTA 720
              |  |     | || |||  |||  |  |    |||| | |  ||  |   |  | |||
Db       2359 ACACCTGAGGTGCCGTTCTTCTCGACACTCGAAGGCGCCTGGATCACCGAGCCGGTGCTC 2418

Qy        721 GACGCGGCGTACTGGTACCGCAACATGTCGAGCACGGTCCGGTTCGAGCCCGCCGCCCGG 780
              ||||   | ||||||||||||||| |       |   ||| | ||||  |||||||  |  |
Db       2419 GACGGCACCTACTGGTACCGCAACCTCCGCCACCGCGTCGGCTTCGCCCCCGCCGTCGAG 2478

Qy        781 ---CTGCTTCTGCAGCAGGGGCCCAAGACGTTCGTCGAGATGAGCCCGCACCCGGTGCTG 837
              ||       |  | ||   ||      ||| ||||| | ||| | ||||| || ||
Db       2479 ACCCTCGCCACCGACGAAGGCTTCACCCACTTCATCGAGGTCAGCGCCCACCCCGTCCTC 2538

Qy        838 ACCATGGGCCTCCAGGAGCTCGCCCCGGACCTGGGCGACACCACCGGC 885
              ||||||  |||||  |||  || | | | ||| |||  |  |  | ||
Db       2539 ACCATGACCCTCCCCGAGACCGTCACCGGCCTCGGCACCCTCCGCCGC 2586
```

Search completed: June 18, 2003, 05:08:04
Job time : 120.124 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 17, 2003, 11:26:45 ; Search time 1215.24 Seconds
(without alignments)
12633.968 Million cell updates/sec

Title: US-09-914-286-1_COPY_85_1032
Perfect score: 948
Sequence: 1 gtcttcgtttttcccgggca.....gcaccctggaccacttcctg 948

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		93	9.8	1075	14	BQ679564	BQ679564 AGENCOURT
c 2		88.8	9.4	1798	17	AG171124	AG171124 Pan trogl
c 3		85.2	9.0	710	17	AQ159669	AQ159669 mgxb0001E
c 4		85.2	9.0	1424	14	BQ720687	BQ720687 AGENCOURT
c 5		84.2	8.9	878	12	BG786339	BG786339 SEAUMC006
c 6		82	8.6	943	12	BG785567	BG785567 SEAUMC005
c 7		81	8.5	932	17	CNS0072Q	AL066742 Drosophil
8		78.6	8.3	1281	12	BG852363	BG852363 1024034A0
c 9		78.2	8.2	1169	17	AG032383	AG032383 Pan trogl
c 10		78.2	8.2	1435	17	AG060835	AG060835 Pan trogl
11		76.4	8.1	842	17	AG058791	AG058791 Pan trogl
12		76	8.0	776	17	CNS010RY	AL099352 Drosophil
c 13		76	8.0	1203	17	CNS015Y4	AL106054 Drosophil
c 14		75.4	8.0	873	12	BG844318	BG844318 1024005H1
15		75	7.9	753	17	CNS077R7	AL433049 T7 end of
c 16		74.8	7.9	1516	12	BG809984	BG809984 mgct002xd
17		74	7.8	659	12	BE725078	BE725078 894080H05
c 18		74	7.8	925	17	CNS0091P	AL053013 Drosophil
19		74	7.8	956	17	AG032985	AG032985 Pan trogl
c 20		74	7.8	1046	14	BQ952554	BQ952554 AGENCOURT
c 21		73.6	7.8	1651	12	BG809816	BG809816 mgct001xk
22		73.4	7.7	1149	14	BM910848	BM910848 AGENCOURT
23		72.6	7.7	1101	17	CNS01523	AL104901 Drosophil
c 24		72.2	7.6	1116	12	BG810038	BG810038 mgct002xi
c 25		72	7.6	1169	14	BQ922913	BQ922913 AGENCOURT
c 26		72	7.6	1190	14	BQ673485	BQ673485 AGENCOURT
27		72	7.6	1218	14	BQ876717	BQ876717 AGENCOURT
28		71.8	7.6	1008	14	BQ942600	BQ942600 AGENCOURT
29		71.8	7.6	1131	17	AG042920	AG042920 Pan trogl
30		71.6	7.6	888	17	AG030591	AG030591 Pan trogl
c 31		71.6	7.6	1126	17	AG064051	AG064051 Pan trogl
32		71.6	7.6	1242	14	BM911414	BM911414 AGENCOURT
33		71.4	7.5	1004	17	AG060160	AG060160 Pan trogl
34		71	7.5	668	12	BG859575	BG859575 1024064B1
35		71	7.5	691	17	AG171171	AG171171 Pan trogl
36		71	7.5	925	17	CNS0091P	AL053013 Drosophil
37		70.8	7.5	1121	17	AG062320	AG062320 Pan trogl
c 38		70.8	7.5	1160	14	BM926731	BM926731 AGENCOURT
c 39		70.8	7.5	1169	13	BM549879	BM549879 AGENCOURT
c 40		70.6	7.4	685	12	BG786336	BG786336 SEAUMC006
41		70.6	7.4	975	17	AG046230	AG046230 Pan trogl
c 42		70.6	7.4	1136	14	BQ943816	BQ943816 AGENCOURT
c 43		70.2	7.4	997	17	AG057392	AG057392 Pan trogl
44		70	7.4	935	17	CNS006XK	AL066051 Drosophil
c 45		70	7.4	1231	17	AG064011	AG064011 Pan trogl

ALIGNMENTS

RESULT 1
BQ679564
LOCUS BQ679564 1075 bp mRNA linear EST 15-JUL-2002
DEFINITION AGENCOURT_8210413 NIH_MGC_112 Homo sapiens cDNA clone IMAGE:6259089 5', mRNA sequence.
ACCESSION BQ679564
VERSION BQ679564.1 GI:21792243
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1075)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: DCTD/DTP
cDNA Library Preparation: Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM2416 row: a column: 10
High quality sequence stop: 252.
FEATURES Location/Qualifiers
source 1. .1075

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:6259089"
/clone_lib="NIH_MGC_112"
/tissue_type="melanotic melanoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: skin; Vector: pOTB7; Site_1: XhoI; Site_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH_MGC Library."

BASE COUNT 47 a 522 c 420 g 51 t 35 others
ORIGIN

Query Match 9.8%; Score 93; DB 14; Length 1075;
Best Local Similarity 44.3%; Pred. No. 2.3e-07;
Matches 358; Conservative 0; Mismatches 448; Indels 3; Gaps 1;

```
Qy        90 CGCCTGCGAAGCCGCGTTCGCGCCCTACGTCGACTGGTCGGTGGAGCAGGTGTTGCGGGA 149
Db       256 CCCCNTCCNGCCGCNNNCCCCCCNCNCGGGCGACTCCGGGCCGGAGCGCCGCCCGGGCGC 315

Qy       150 CTCGCCGGACGCTCCCGGGCTGGACCGGGTGGACGTCGTCCAGCCGACCCTGTTCGCCGT 209
Db       316 GGCCGCCGGGGGCGGCGGGCGGCCCCCGGGGGCGCCGGCGCGCGCGGGCCGGCCCGCGCG 375

Qy       210 CATGATCTCCCTGGCCGCCCTCTGGCGCTCGCAAGGGGTCGAGCCGTGCGCGGTGCTGGG 269
Db       376 GCGGCGCCCCGGCGCGGGGGCGGGGCGGCCGNGCGCGGCCCCGCGGCCCGGGCCCCGGGG 435

Qy       270 ACACAGCCTGGGCGAGATCGCGGCAGCCCACGTCTCGGGAGGCCTGTCCCTGGCCGACGC 329
Db       436 GCCCGGGGGGCCCGGGGGCGGGGCCGCCCGCGGGCGGGCGGGCCGGCGGCGCGGGCCCGC 495

Qy       330 CGCACGCGTGGTGACGCTTTGGAGCCAGGCACAGACCACCCTTGCCGGGACCGGCGCGCT 389
Db       496 CCGCGGCCGCCCGGCNGCCCGCCGCCGGGGCCNCCCTCCCCGCCGCGCGGGCGCCCCCCC 555

Qy       390 CGTCTCCGTCGCCGCCACGCCGGATGAGCTCCTGCCCCGAATCGCTCCGTGGACCGAGGA 449
Db       556 CGGCNCCCCGGCCGCCCCGCCCGGCGNGCCCCCCCCCCCCCCCCGCGGGGGGCCGGGCG 615

Qy       450 CAACCCGGCGCGGCTCGCCGTCGCAGCCGTCAACGGACCCCGGAGCACAGTCGTTTCCGG 509
Db       616 GGCCCGCCCGCCCCGGGGGCCCGCCCCGGCGGCGCCCCGCCGCGCGCCGCGGGGGCGGC 675

Qy       510 TGCCCGCGAGGCCGTCGCGGACCTGGTGGCCGACCTCACCGCCGCGCAGGTGCGCACGCG 569
Db       676 CGCGCCCGCGCGCGGGGCCGGCCGGGGGGCGGGGCGCGCGGGGCGCCGCGCGCGGCCGCC 735

Qy       570 CATGATCCCGGTGGACGTTCCCGCCCACTCCCCCCTGATGTACGCCATCGAGGAACGGGT 629
Db       736 CGGGGCCCGGCCCNCCG---CCGCGCCCCCCCCCCCCCCGCCCCCGCCCCGGCGGNCCGC 792

Qy       630 CGTCAGCGGCCTGCTGCCCATCACCCCACGCCCCTCCCGCATCCCCTTCCACTCCTCGGT 689
Db       793 CGCCCCGCGCCCGCGCCCGGNNGCGCCCCGGGCCGCCCGNNCGGGCCCCGGGCGGCCGGC 852

Qy       690 GACCGGCGGCCGCCTCGACACCCGCGAGCTAGACGCGGCGTACTGGTACCGCAACATGTC 749
Db       853 CCCCCCCCGGCGCCCCGCCCGCCCCCCGGNGCCCCGCGCCCGGCGCCGCCGCCCGCGGCC 912

Qy       750 GAGCACGGTCCGGTTCGAGCCCGCCGCCCGGCTGCTTCTGCAGCAGGGGCCCAAGACGTT 809
Db       913 CCGCGCCGNCCGCGGCGGGGGCCCCGCNNGGCCGCGGCGGCCGGNGGCGCGGCGNCGGGG 972

Qy       810 CGTCGAGATGAGCCCGCACCCGGTGCTGACCATGGGCCTCCAGGAGCTCGCCCCGGACCT 869
Db       973 CGCCGNCGCCGCCCCGCGGCGGCCNGCCGNCGCGCCCCCCCGCCGGGCGGGGCNGGCGCG 1032
```

Qy 870 GGGCGACACCACCGGCACCGCCGACACCG 898
Db 1033 CCGCGGGCGCGGGGGCCGCGGCCGCGCGG 1061

RESULT 2
AG171124/c
LOCUS AG171124 1798 bp DNA linear GSS 09-JAN-2002
DEFINITION Pan troglodytes DNA, clone: RP43-040F09.TJ, genomic survey sequence.
ACCESSION AG171124
VERSION AG171124.1 GI:16700802
KEYWORDS GSS.
SOURCE Pan troglodytes male lymphocytes DNA, clone_lib:RPCI-43 Chimpanzee Male BAC Library clone:RP43-040F09.TJ.
ORGANISM Pan troglodytes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE 1
AUTHORS Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE BAC end sequences of Library RPCI-43
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1798)
AUTHORS Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE Direct Submission
JOURNAL Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC); 1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:chimpbes@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/, Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT Clones are derived from the chimpanzee BAC library RPCI-43 This BAC end was generated during the R&D process and may have higher chance of clone tracking errors.
PRIMERS
Sequencing: TJ
LIBRARY
Vector : pBACe3.6
R.Site 1 : EcoRI
R.Site 2 : EcoRI.
FEATURES Location/Qualifiers
source 1. .1798
/organism="Pan troglodytes"
/db_xref="taxon:9598"
/clone="RP43-040F09.TJ"
/sex="male"
/cell_type="lymphocytes"
/clone_lib="RPCI-43 Chimpanzee Male BAC Library"
BASE COUNT 0 a 778 c 655 g 0 t 365 others
ORIGIN

Query Match 9.4%; Score 88.8; DB 17; Length 1798;
Best Local Similarity 39.8%; Pred. No. 1.3e-06;
Matches 367; Conservative 0; Mismatches 544; Indels 10; Gaps 3;

```
Qy        15 CGGGCAGGGCCCGCAATGGCCGGGCATGGGAAGGGAACTTCTCGACGCTTCCGACGTCTT 74
Db      1135 CGNGNGGCGGCCCGCGGNCGCGCGCGCGGGCGGGCGNCGCGGNGNGNCGCGGGCGCGNNN 1076

Qy        75 CCGGGAGAGCGTCCGCGCCTGCGAAGCCGCGTTCGCGCCCTACGTCGACTGGTCGGTGGA 134
Db      1075 GCGCGCGCGCGCNCGCGCGCGCGCGCCGGCGCNNNNNGGCGCGCGGGCGNGCNGCNGGC 1016

Qy       135 GCAGGTGTTGCGGGACTCGCCGGACGCTCCCGGGCTGGACCGGGTGGACGTCGTCCAGCC 194
Db      1015 GCGCCGNGGGCGGCGGCGGCGCGCGGGCGGCGNGCGCCGCGGCGGCGNGNGGCGCGGGGC 956

Qy       195 GACCCTGTTCGCCGTCATGATCTCCCTGGCCGCCCTCTGGCGCTCGCAAGGGGTCGAGCC 254
Db       955 GGCGNNGNGCGCCGNGCGGGCGNGGCGGGCGCGGCGCCGCGGNNCGCGCGGGCGCGNGGC 896

Qy       255 GTGCGCGGTGCTGGGACACAGCCTGGGCGAGATCGCGGCAGCCCACGTCTCGGGAGGCCT 314
```

```
Db        895 GCGCGGCGCGGCG------NGCGCGCGCGGCGNCGCGGGCGNGCGCGCCGCGGGGGCGNC 842

Qy        315 GTCCCTGGCCGACGCCGCACGCGTGGTGACGCTTTGGAGCCAGGCACAGACCACCCTTGC 374

Db        841 GCGCGGCGCGNGCGCGGCGCGCNNNGCGCCGCGGCGGCGCGCGGNGNGCGGGCGCCGCGG 782

Qy        375 CGGGACCGGCGCGCTCGTCTCCGTCGCCGCCACGCCGGATGAGCTCCTGCCCCGAATCGC 434

Db        781 CGGGCGNCCGGCGCGCGCGNGCGCGNCGCGCGCGCGCGCGCCGCGCCGCGCGCG---CGC 725

Qy        435 TCCGTGGACCGAGGACAACCCGGCGCGGCTCGCCGTCGCAGCCGTCAACGGACCCCGGAG 494

Db        724 GCGCGCGCGCGCGCGCGCGCGCGCGCNCGNCCGCGCGCGCCGCGGGCGCCCGCGGCCGCGC 665

Qy        495 CACAGTCGTTTCCGGTGCCCGCGAGGCCGTCGCGGACCTGGTGGCCGACCTCACCGCCGC 554

Db        664 CCGCGCCGCGCGCGGCGCCCGCGCNCCCCGCGCGCCGCGCGCGCCCGCGCGCGCCGCCNC 605

Qy        555 GCAGGTGCGCACGCGCATGATCCCGGTGGACGTTCCCGCCCACTCCCCCCTGATGTACGC 614

Db        604 CCCNNNCCCCCCCCCCCCNCGCNCGCCCCNCCCCCCCNNNCGCGCCNCCCCGCGNCCCNC 545

Qy        615 CATCGAGGAACGGGTCGTCAGCGGCCTGCTGCCCATCACCCCACGCCCCTCCCGCATCCC 674

Db        544 CNNNCCNCNCNNNNCCCCCGCCCNCCCCCGCGCCCCCNNNNNNCCGNNNNNCGCCNCCCC 485

Qy        675 CTTCCACTCCTCGGTGACCGGCGGCCGCCTCGA-CACCCGCGAGCTAGACGCGGCGTACT 733

Db        484 CCNCCCCCNNNCCCCNNCCGGGGGCGGGGGGGCGCGGCGGGGCGGCGNGCGCGGGGGGCG 425

Qy        734 GGTACCGCAACATGTCGAGCACGGTCCGGTTCGAGCCCGCCGCCCCGGCTGCTTCTGCAGC 793

Db        424 GCCGGCGCGGCGNGGCGCGGCCGGCGGGGNGCCGNNGCGGNGGCGGGCGGCCGGCGCCGC 365

Qy        794 AGGGGCCCAAGACGTTCGTCGAGATGAGCCCGCACCCGGTGCTGACCATGGGCCTCCAGG 853

Db        364 GGCGGCGGCGGGCGGGGGCCGGNGCGGNGGCGGCGGNNNNNNNNNNNNNNNNNNGGNGNN 305

Qy        854 AGCTCGCCCCGGACCTGGGCGACACCACCGGCACCGCCGACACCGTGATCATGGGCACGC 913

Db        304 NNNNNNNNNNNNNNNNNNNNNGNNCCGCCCCGCGGNNNNGGGCCGCCGCGGCGGCCGCG 245

Qy        914 TGCGCCGCGGCCAGGGCACCC 934

Db        244 CGCGCCGCGGCCGCNGCGCGC 224


RESULT 3
AQ159669/c
LOCUS       AQ159669                710 bp    DNA     linear   GSS 09-SEP-1998
DEFINITION  mgxb0001E02r CUGI Rice Blast BAC Library Magnaporthe grisea genomic
            clone mgxb0001E02r, DNA sequence.
ACCESSION   AQ159669
VERSION     AQ159669.1  GI:3556694
KEYWORDS    GSS.
SOURCE      Magnaporthe grisea.
  ORGANISM  Magnaporthe grisea
            Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
            Sordariomycetes incertae sedis; Magnaporthaceae; Magnaporthe.
REFERENCE   1  (bases 1 to 710)
  AUTHORS   Yu,Y., Zhu,H., Boyd,C.A., Gaudette,B., Gayle,A., Kingsbury,R.,
            Phillips,K., Sasinowski,M, Wing,R.A. and Dean,R.A.
  TITLE     A BAC End Sequencing Framework to Sequence the Magnaporthe grisea
            Genome
  JOURNAL   Unpublished (1998)
COMMENT     Contact: Dean RA
            Clemson University Genomics Institute
            Clemson University
            100 Jordan Hall, Clemson Universiy, Clemson, SC 29634
            Tel: 864 656 5737
            Fax: 864 656 4293
            Email: rdean@clemson.edu
            Seq primer: GGAAACAGCTATGACCATG
            Class: BAC ends
            High quality sequence stop: 359.
FEATURES             Location/Qualifiers
     source          1. .710
                     /organism="Magnaporthe grisea"
                     /strain="70-15"
                     /db_xref="taxon:148305"
                     /clone="mgxb0001E02r"
                     /clone_lib="CUGI Rice Blast BAC Library"
                     /tissue_type="Protoplasts"
                     /lab_host="E. coli DH10B"
                     /note="Vector: pBACWICH; Site_1: HindIII; Site_2: HindIII;
                     Rice blast is one of the most devestating fungal diseases
                     of rice  world wide.  It is a filamentous ascomycete with
                     a haploid genome (n=7)  of approximately 40 Mbp.  Rice
                     blast is an important model fungal pathogen  for studying
                     numerous aspects of the fungal-host interaction.   In
                     order to facilitate genome wide analysis, a BAC library
                     containing  9216 clones with an average insert size of 130
                     kbp was constructed.   This library represents greater
                     than 25X genome coverage.  High density  colony filters
                     are available upon request."
BASE COUNT      156 a    185 c    242 g    126 t      1 others
ORIGIN

  Query Match             9.0%;  Score 85.2;  DB 17;  Length 710;
  Best Local Similarity   53.6%;  Pred. No. 5.8e-06;
  Matches  177;  Conservative    0;  Mismatches  153;  Indels    0;  Gaps    0;

Qy          1 GTCTTCGTTTTTCCCGGGCAGGGCCCGCAATGGCCGGGCATGGGAAGGGAACTTCTCGAC 60

Db        353 GCCTTCATGTTCACTGGCCAGGGCTCTCACTACGCTGGTCTTGGAAAGGAGCTGTACGCC 294

Qy         61 GCTTCCGACGTCTTCCGGGAGAGCGTCCGCGCCTGCGAAGCCGCGTTCGCGCCCTACGTC 120

Db        293 AACCACGCTGTCTTCCGCCAGGCCATCAACGAGTTCGACAGGATCTCGCAGATCCACGGC 234

Qy        121 GACTGGTCGGTGGAGCAGGTGTTGCGGGACTCGCCGGACGCTCCCGGGCTGGACCGGGTG 180

Db        233 TTCCCGAGCTTCATGCCTCTGGTCGACGGCAGCGAGCCTGATGTCGCCAAGCTCGAGCCC 174

Qy        181 GACGTCGTCCAGCCGACCCTGTTCGCCGTCATGATCTCCCTGGCCGCCCTCTGGCGCTCG 240

Db        173 CTCATCGTCCAGCTTGGTCTCTGCTGCTTCGAGATGGCCTTGGCCAAGCTCTGGGCTTCT 114

Qy        241 CAAGGGGTCGAGCCGTGCGCGGTGCTGGGACACAGCCTGGGCGAGATCGCGGCAGCCCAC 300

Db        113 TGGGGTGTCAAGCCTGCTGCGGTCATGGGCCACAGCCTGGGCGAGTACGCTGCTCTCCAC 54

Qy        301 GTCTCGGGAGGCCTGTCCCTGGCCGACGCC 330

Db         53 GTTGCCGGTGTTCTGTCTGCCAGCGACACC 24


RESULT 4
BQ720687/c
LOCUS       BQ720687                1424 bp    mRNA    linear   EST 16-JUL-2002
DEFINITION  AGENCOURT_8234644 Lupski_sympathetic_trunk Homo sapiens cDNA clone
            IMAGE:6188776 5', mRNA sequence.
ACCESSION   BQ720687
VERSION     BQ720687.1  GI:21859584
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1424)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
```

```
          Tissue Procurement: Dr. James R. Lupski
           cDNA Library Preparation: Life Technologies, Inc.
           cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
           DNA Sequencing by: Agencourt Bioscience Corporation
           Clone distribution: MGC clone distribution information can be
          found through the I.M.A.G.E. Consortium/LLNL at:
          http://image.llnl.gov
          Plate: LLAM13584  row: o  column: 17
          High quality sequence stop: 243.
FEATURES             Location/Qualifiers
     source          1. .1424
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:6188776"
                     /clone_lib="Lupski_sympathetic_trunk"
                     /sex="male"
                     /tissue_type="sympathetic trunk"
                     /dev_stage="adult, 16 yr"
                     /lab_host="DH10B"
                     /note="Vector: pCMV-SPORT6 (Life Technologies); Site_1:
                     NotI; Site_2: SalI; cDNA made by oligo-dT priming.
                     Directionally cloned using the following adaptors:
                     5'-TCGACCCACGCGTCCG-3' and
                     5'-GACTAGTTCTAGATCGCGAGCGGCCGCCCT(15)-3'. Size selected >
                     1 kb for average insert length 1.9 kb. This is a primary
                     library, non-amplified. Library constructed by Life
                     Technologies and donated  by J. Lupski, M.D./Ph.D. (Baylor
                     College of Medicine); available through Life
                     Technologies."
BASE COUNT       74 a    157 c    766 g     41 t    386 others
ORIGIN

  Query Match             9.0%;  Score 85.2;  DB 14;  Length 1424;
  Best Local Similarity   32.5%;  Pred. No. 5.9e-06;
  Matches  261;  Conservative    0;  Mismatches  542;  Indels    0;  Gaps    0;

Qy        144 GCGGGACTCGCCGGACGCTCCCGGGCTGGACCGGGTGGACGTCGTCCAGCCGACCCTGTT 203
              ||| | | | |  | ||       ||    |  | | |  ||     |  ||    || |
Db       1416 GCGCGCCGCNCGCGGCGGNGNGCGGGCNGCGGGCGGGCCCGGGCGCNNGCGNCGCCCGGC 1357

Qy        204 CGCCGTCATGATCTCCCTGGCCGCCCTCTGGCGCTCGCAAGGGGTCGAGCCGTGCGCGGT 263
              ||       |  | | |  ||||| | | | | | ||   |||| |  || | ||   |
Db       1356 CGNGCGGGCGCGCNCGCCCGCCGCGCGCCGCCCCGCGNNCGGGGGCNCGCGGCGCCGNGC 1297

Qy        264 GCTGGGACACAGCCTGGGCGAGATCGCGGCAGCCCACGTCTCGGGAGGCCTGTCCCTGGC 323
              |  ||  | |   | || |  |  |    |  ||| ||   ||    | | |  ||  ||
Db       1296 GNGGGNGCGCGCGCGGGNCCGGCGCCGNCCCNCCCCCGGNCCGCNCCGGCGGGNCCGCGC 1237

Qy        324 CGACGCCGCACGCGTGGTGACGCTTTGGAGCCAGGCACAGACCACCCTTGCCGGGACCGG 383
              | |||   ||   |  |       |  ||   |  | | ||  |              |
Db       1236 GGGGGCCCGCCGGCGGCGGNGCGCGCGCNGCNCCGGCCCGCCCNCNNNNNNNCNCNNCCC 1177

Qy        384 CGCGCTCGTCTCCGTCGCCGCCACGCCGGATGAGCTCCTGCCCCGAATCGCTCCGTGGAC 443
              ||    |  | |  |    |  | |       ||       |      |
Db       1176 CGNCNNCCCCNCNNNCNNNNCGGCNCNCNCNNCNCNCNNCNGGCNNCCNCNNNCNCNNCN 1117

Qy        444 CGAGGACAACCCGGCGCGGCTCGCCGTCGCAGCCGTCAACGGACCCCGGAGCACAGTCGT 503
              |       | |    | |  |  |    |       |     ||    | |
Db       1116 CNCNCNNCNCNCCNNNCNGGCCNNCCNNNCCNNNNNNCCCNNCNNCCCNCNCNCNNNNNC 1057

Qy        504 TTCCGGTGCCCGCGAGGCCGTCGCGGACCTGGTGGCCGACCTCACCGCCGCGCAGGTGCG 563
              |       |  |    ||  | |          ||    |  ||| |       |
Db       1056 CCCNCNNNNCNNCNCNNCCCCCCCCCCNNNCNNNNCNNCCCCCNNCNCCCNCNNNCNNCCN 997

Qy        564 CACGCGCATGATCCCGGTGGACGTTCCCGCCCACTCCCCCCTGATGTACGCCATCGAGGA 623
              | | |        ||          | |   | | ||||||        | ||  |
Db        996 CNCNCNNCCCNCNCCNNCCNNNCNNCNCCNNCCCCCCCCCCCCNNNCCCCCCCCCNCNCC 937

Qy        624 ACGGGTCGTCAGCGGCCTGCTGCCCATCACCCCACGCCCCTCCCGCATCCCCTTCCACTC 683
              |       |     ||     | |    |||| | |||| |||    |||| || | |
Db        936 NCCCCNNNNCNNNNNCCNNNNNCNCNCNCCCCCNCNCCCCCCCCNNCCCCCCCCCCCCCC 877

Qy        684 CTCGGTGACCGGCGGCCGCCTCGACACCCGCGAGCTAGACGCGGCGTACTGGTACCGCAA 743
```

```
                  ||  ||||| ||    | |||                    |     ||
Db        876 NCNNCNNNCCCCCGGCCCCCCNNCCNCCCCNCCNNNCCNNNCNNCCNCCCNCCCCCNCC 817

Qy        744 CATGTCGAGCACGGTCCGGTTCGAGCCCGCCGCCCGGCTGCTTCTGCAGCAGGGGCCCAA 803
              |         |      |    | | || |||      |     |  |     |
Db        816 CNNNNNNNCCNNNNCCNCCCCCCCNCNCNCCNCCCCCNCCCNCGCNCCNCNCNCCCNNCC 757

Qy        804 GACGTTCGTCGAGATGAGCCCGCACCCGGTGCTGACCATGGGCCTCCAGGAGCTCGCCCC 863
                |  |           |||    ||        |      || ||     | | |||
Db        756 CCCCNNCCNNCCCCCCCCCCCNNCNCCCNCCNNCCCNNCCCCCCCCCNCNCCCCCCCCCN 697

Qy        864 GGACCTGGGCGACACCACCGGCACCGCCGACACCGTGATCATGGGCACGCTGCGCCGCGG 923
                ||    |  | || ||  | || ||  |         |      | | |  | || |
Db        696 NCCCCCNCCCCCCNCCCCCCCCCCCCCCCNNCNNNCNCCCCCCNNCCCCCCCCCNCCCCCN 637

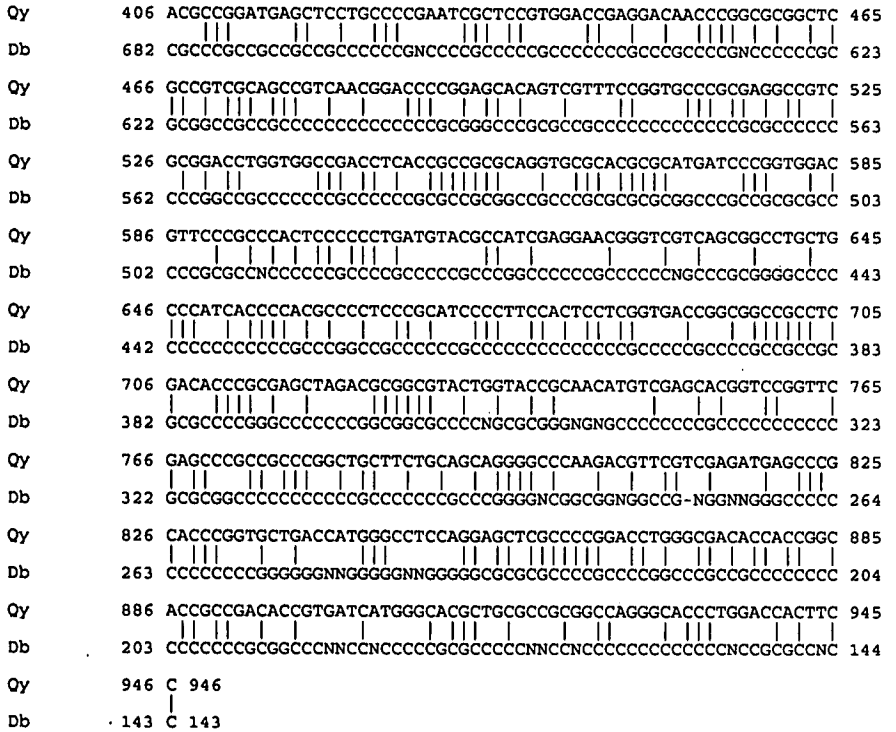
Qy        924 CCAGGGCACCCTGGACCACTTCC 946
                |   | |||     | |  ||
Db        636 NCCCNCCCCCCCCCNNCCCCNCC 614


RESULT 5
BG786339/c
LOCUS       BG786339                878 bp    mRNA    linear   EST 20-MAY-2001
DEFINITION  SEAUMC006296 Sea urchin primary mesenchyme cell cDNA library
            Strongylocentrotus purpuratus cDNA clone PC_0028_A2_H08_MR 5', mRNA
            sequence.
ACCESSION   BG786339
VERSION     BG786339.1  GI:14157352
KEYWORDS    EST.
SOURCE      Strongylocentrotus purpuratus.
  ORGANISM  Strongylocentrotus purpuratus
            Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
            Echinoidea; Euechinoidea; Echinacea; Echinoida;
            Strongylocentrotidae; Strongylocentrotus.
REFERENCE   1  (bases 1 to 878)
  AUTHORS   Zhu,X., Mahairas,G., Illies,M.R., Cameron,R.A., Davidson,E.H. and
            Ettensohn,C.A.
  TITLE     A large scale analysis of mRNAs expressed by primary mesenchyme
            cells of the sea urchin embryo
  JOURNAL   Development 128 (13), 2615-2627 (2001)
  MEDLINE   21384984
COMMENT     Contact: Ettensohn CA
            Dept. Biol. Sci.
            Carnegie Mellon University
            4400 Fifth Avenue, Pittsburgh, PA 15213, USA
            Tel: +1 412 268 5849
            Email: ettensohn@andrew.cmu.edu.
FEATURES             Location/Qualifiers
     source          1. .878
                     /organism="Strongylocentrotus purpuratus"
                     /db_xref="taxon:7668"
                     /clone="PC_0028_A2_H08_MR"
                     /clone_lib="Sea urchin primary mesenchyme cell cDNA
                     library"
                     /tissue_type="embryo"
                     /cell_type="primary mesenchyme cells"
                     /lab_host="E.coli"
                     /note="Vector: pSPORT1; Site_1: NotI; Site_2: SalI; oligo
                     dt priming from poly A+ RNA, directionally cloned"
BASE COUNT       22 a    227 c    572 g      8 t     49 others
ORIGIN

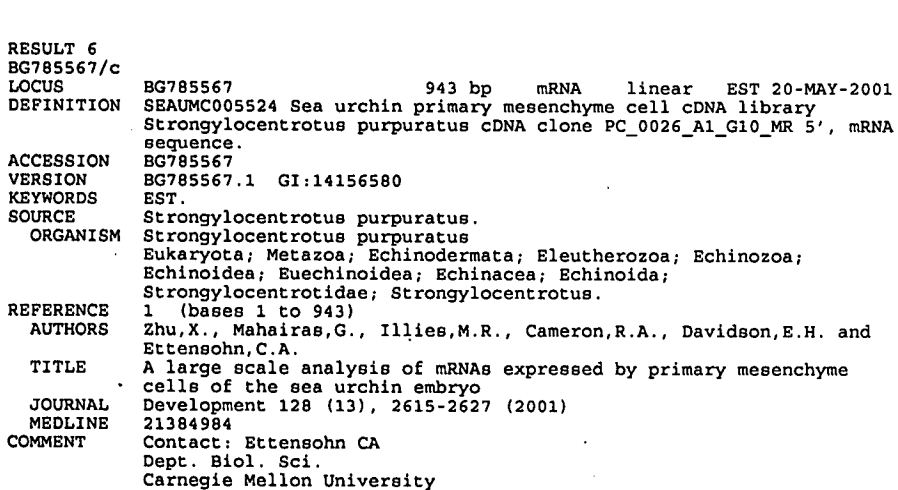
  Query Match             8.9%;  Score 84.2;  DB 12;  Length 878;
  Best Local Similarity   46.0%;  Pred. No. 8.8e-06;
  Matches  304;  Conservative    0;  Mismatches  354;  Indels    3;  Gaps    2;

Qy        288 CGCGGCAGCCCACGTCTCGGGAGGCCTGTCCCTGGCCGACGCCGCACGCGTGGTGACGCT 347
              ||||  |  |||       | || | ||   ||| | ||  | |  | |||  || | |||
Db        802 CGCGCCNCCCCGGCGGGCCGGCGCCCCCCCCCGCCCCGCCCGCCGCGCCGGGGGCCGCC 743

Qy        348 TTGGAGCCAGGCACAGACCACCCTTGCCGGGACCGGCGCGCTCGTCTC--CGTCGCCGCC 405
                ||  ||  || |   || |||  ||||   ||  | | |     | |  |  | ||| |
Db        742 CCGGCCCCCCGCGCGCCCCCCCCCCGCCGCCCCCCCCCCCCGNCGCGCGNCCCCCCCGGC 683
```



```
RESULT 6
BG785567/c
LOCUS       BG785567                 943 bp    mRNA    linear   EST 20-MAY-2001
DEFINITION  SEAUMC005524 Sea urchin primary mesenchyme cell cDNA library
            Strongylocentrotus purpuratus cDNA clone PC_0026_A1_G10_MR 5', mRNA
            sequence.
ACCESSION   BG785567
VERSION     BG785567.1  GI:14156580
KEYWORDS    EST.
SOURCE      Strongylocentrotus purpuratus.
  ORGANISM  Strongylocentrotus purpuratus
            Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
            Echinoidea; Euechinoidea; Echinacea; Echinoida;
            Strongylocentrotidae; Strongylocentrotus.
REFERENCE   1  (bases 1 to 943)
  AUTHORS   Zhu,X., Mahairas,G., Illies,M.R., Cameron,R.A., Davidson,E.H. and
            Ettensohn,C.A.
  TITLE     A large scale analysis of mRNAs expressed by primary mesenchyme
            cells of the sea urchin embryo
  JOURNAL   Development 128 (13), 2615-2627 (2001)
  MEDLINE   21384984
COMMENT     Contact: Ettensohn CA
            Dept. Biol. Sci.
            Carnegie Mellon University
            4400 Fifth Avenue, Pittsburgh, PA 15213, USA
            Tel: +1 412 268 5849
            Email: ettensohn@andrew.cmu.edu.
FEATURES             Location/Qualifiers
     source          1. .943
                     /organism="Strongylocentrotus purpuratus"
                     /db_xref="taxon:7668"
                     /clone="PC_0026_A1_G10_MR"
                     /clone_lib="Sea urchin primary mesenchyme cell cDNA
                     library"
                     /tissue_type="embryo"
                     /cell_type="primary mesenchyme cells"
                     /lab_host="E.coli"
                     /note="Vector: pSPORT1; Site_1: NotI; Site_2: SalI; oligo
                     dt priming from poly A+ RNA, directionally cloned"
BASE COUNT       15 a    321 c    540 g     53 t     14 others
ORIGIN
```



```
RESULT 7
CNS0072Q/c
LOCUS       CNS0072Q                 932 bp    DNA     linear   GSS 03-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence T7 end of BAC #
            BACR14B09 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL066742
VERSION     AL066742.1  GI:4945205
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
```

```
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 932)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1. .932
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone="BACR14B09"
                     /clone_lib="RPCI-98"
                     /note="end : T7"
BASE COUNT      155 a    202 c    241 g     91 t    243 others
ORIGIN

  Query Match             8.5%;  Score 81;  DB 17;  Length 932;
  Best Local Similarity   33.0%;  Pred. No. 3.3e-05;
  Matches  153;  Conservative   98;  Mismatches  212;  Indels    0;  Gaps    0;

Qy        109 GCGCCCTACGTCGACTGGTCGGTGGAGCAGGTGTTGCGGGACTCGCCGGACGCTCCCGGG 168
              |:||| |      ||    : ::: || :|  : |  |:  | | ::|: | |||  |||:|
Db        932 GSGCGCGCGCCCGSGGCSSSSSCGGGSCGCSSGCSGSCCGCCGSSCSCGCCGCGGCCGSG 873

Qy        169 CTGGACCGGGTGGACGTCGTCCAGCCGACCCTGTTCGCCGTCATGATCTCCCTGGCCGCC 228
                || |:||  :  || :|:|  |||| ::  |  |:|::    :  :  |: | :| | |
Db        872 GCGGSCSGGCCSCGCGSSGSCGCGCCGCSSGGGGCCSCSSGGCSSCCSSGCSSGCSCCCG 813

Qy        229 CTCTGGCGCTCGCAAGGGGTCGAGCCGTGCGCGGTGCTGGGACACAGCCTGGGCGAGATC 288
              : | :::   : :  : |  :: |:|   ||: | :|:|:  | : ||: |  :  |  :
Db        812 SSCSSSSCGCSCSSCSCGCSSSCGSCCGCCGSCGCSCCGSCGCGSCGCSSGCCSCCGCSS 753

Qy        289 GCGGCAGCCCACGTCTCGGGAGGCCTGTCCCTGGCCGACGCCGCACGCGTGGTGACGCTT 348
              ::|   :    :|   :|:| |: | :  || |: ||  |:|   |||| || | :||
Db        752 SSGCGSSGGGGSGGGGSGSGGGSGCGSSGCCSGSGCGCGGSCCGSCGCGCGGGGGSGCSS 693

Qy        349 TGGAGCCAGGCACAGACCACCCTTGCCGGGACCGGCGCGCTCGTCTCCGTCGCCGCCACG 408
               ||  :   |:      | |:|   |::|  :|::   || ||, | |   :|:    : :
Db        692 GGGGCSGGCGSSGCCGGCSCSCCSCCSSGCCSCSSGCGGCGCGCCSCGCGSGSGCGSSSC 633

Qy        409 CCGGATGAGCTCCTGCCCCGAATCGCTCCGTGGACCGAGGACAACCCGGCGCGGCTCGCC 468
              |||       | :|  ||||    |:  :|:    ||   :| :|:::: | | :| : :|
Db        632 CCGCGCCSCCGSCGCCCCCYGSCCSGGSCSSCCGCCMCMVARMAVSVSCCCCCSCCMASC 573

Qy        469 GTCGCAGCCGTCAACGGACCCCGGAGCACAGTCGTTTCCGGTGCCCGCGAGGCCGTCGCG 528
              |||:: |: |  : : ||||| |:|:||| |    |:|   |:| ||   :|| | :
Db        572 CCCGCVSGCSCCSCSMSCCCCCGCASCVCAGCCMSAGCMGCGCCMCCCGCCASCGSCCSC 513

Qy        529 GACCTGGTGGCCGACCTCACCGCCGCGCAGGTGCGCACGCGCA 571
              :| :    ||  ||   ::||| | |:|:  | : ||:|::
Db        512 CCYCGBCCCCCCCGCCMGCMVGCCMCCCMGVMACCMGCGVGMM 470


RESULT 8
BG852363
LOCUS       BG852363               1281 bp    mRNA    linear   EST 29-MAY-2001
DEFINITION  1024034A03.y2 C. reinhardtii CC-1690, normalized, Lambda Zap II
            Chlamydomonas reinhardtii cDNA, mRNA sequence.
ACCESSION   BG852363
VERSION     BG852363.1  GI:14233547
KEYWORDS    EST.
SOURCE      Chlamydomonas reinhardtii.
  ORGANISM  Chlamydomonas reinhardtii
            Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
            Chlamydomonadaceae; Chlamydomonas.
REFERENCE   1  (bases 1 to 1281)
  AUTHORS   Grossman,A., Davies,J., Federspiel,N., Harris,E., Lefebvre,P.,
            McDermott,J.P., Silflow,C., Stern,D. and Surzycki,R.
  TITLE     Analyses of the Chlamydomonas reinhardtii Genome: A Model,
            Unicellular System for Analyzing Gene Function and Regulation in
            Vascular Plants; project phase 2
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Charles Hauser
            DCMB Box 91000
            Duke University
            Durham, NC 27708-1000
            Tel: 919 613 8159
            Fax: 919 613 8177
            Email: chauser@duke.edu.
FEATURES             Location/Qualifiers
     source          1. .1281
                     /organism="Chlamydomonas reinhardtii"
                     /strain="CC-1690 wild type mt+ 21gr"
                     /db_xref="taxon:3055"
                     /clone_lib="C. reinhardtii CC-1690, normalized, Lambda Zap
                     II"
                     /note="Vector: pBluescript II SK-; Site_1: EcoRI; Site_2:
                     XhoI; This library, constructed by John Davies and Jeffrey
                     McDermott, combines cDNAs from CC-1690 cells grown to
                     mid-log phase in TAP (acetate-containing) medium in the
                     light, TAP medium in the dark, HS (minimal) medium in
                     ambient levels of CO2 and HS medium bubbled with 5% CO2.
                     PolyA mRNA was purified from each sample, pooled and cDNA
                     synthesized. The cDNA was directionally cloned into lambda
                     ZAP II (Stratagene) in the EcoRI (5') and XhoI (3') sites.
                     pBluescript II SK- plasmids were excised from the lambda
                     ZAP clones by superinfection with ExAssist (Stratagene)
                     phage. The library was normalized using method 4 described
                     in Bonaldo et al (1996) Genome Research 6: 791-806."
BASE COUNT       30 a    697 c    456 g     15 t     83 others
ORIGIN

  Query Match             8.3%;  Score 78.6;  DB 12;  Length 1281;
  Best Local Similarity   44.3%;  Pred. No. 9e-05;
  Matches  285;  Conservative    0;  Mismatches  356;  Indels    2;  Gaps    1;

Qy        144 GCGGGACTCGCCGGACGCTCCCGGGCTGGACCGGGTGGACGTCGTCCAGCCGACCCTGTT 203
              || ||   |||| | ||| | |      |  |·|      |  || |  |||| |||
Db        435 GCCGGCGGCGCCCGGCGCNCGCCCCNCCGGGCCGCGNCCCCCCGCCGGGCCGGCCCCCNG 494

Qy        204 CGCCGTCATGATCTCCCTGGCCGCCCTCTGGCGCTCGCAAGGGGTCGAGCCGTGCGCGGT 263
                  |  |  |    | | | |  |  | || | |||       ||  |||  | |
Db        495 GCNCCCCNGGCNGCCGCGGCCGCGCGCCGCGCCCGCGCCCCCCCCNCGCCCCGGCCCCCCG 554

Qy        264 GCTGGGACACAGCCTGGGCGAGATCGCGGCAGCCCACGTCTCGGGAGGCCTGTCCCTGGC 323
              | | | | |  |  |||| |  |||  |  ||  |  | |      ||| | | | |
Db        555 CCCGCGGCCCGCCNCCGGCGCGCGCGCCCCGCCCGGCCCCCCNNCCCGCCCGGCGCGGCG 614

Qy        324 CGACGCCGCACGCGTGGTGACGCTTTGGAGCCAGGCACAGACCACCCTTGCCGGGACCGG 383
              || || ||  |||| ||   | |   |   |   |   |  || |   |      | ||
Db        615 CGCCGGCGNGCGCGCGGNCCCCCCCCGCCGGGCGCGGCCCCCCCCCGGGGGGGNCCGCGGG 674

Qy        384 CGCGCTCGTCTCCGTCGCCGCCACGCCGGATGAGCTCCTGCCCCGAATCGCTCCGTGGAC 443
              ||||   |    | | || || |||| |    | ||  ||||   | | |  || |
Db        675 CGCGGCGGCGCGCCCCCCCCCCNCGCCCGCCCGCCGCCCCCCCCCGCCCCCGCGCNGGCC 734

Qy        444 CGAGGACAACCCGGCGCGGCTCGCCGTCGCAGCCGTCAACGGACCCCGGAGCACAGTCGT 503
              || | | ||||  |  | ||    | | | |      |  || |  || | | ||
Db        735 CGCGCGCCCCCCGCGGNCNCCCGGGCGCNCCCCGGCGGGGCGCGCCGGCCGCGCCGGCCG 794
```

```
Qy        504 TTCCGGTGCCCGCGAGGCCGTCGCGGACCTGGTGGCCGACCTCACCGCCGCGCGCAGGTGCG 563
              ||| | || ||  |||| |||    | ||   ||  |  | |  || ||    |  ||
Db        795 NCCCGCCGGCCCCGGCGCCGGCGCCNCGCGGGCCCCCCCCCGCCNCGCCCCCGGCCGGCCG 854

Qy        564 CACGCGCATGATCCCGGTGGACGTTCCCGCCCACTCCCCCCTGATGTACGCCATCGAGGA 623
              | |||      || ||  |      || | |    | ||              ||||   | |
Db        855 GCCCCGCCGCGCCCGGGCCGGGCGCCCGGGCGGGGCGCGGGCCNGCGCCGCCGCGGCGCG 914

Qy        624 ACGGGTCGTCAGCGGCCTGCTGCCCATCACCCCACGCCCCTCCCGCATCCCCTTCCACTC 683
              | |       | |||| ||  ||   | | || | | | || ||  ||   |  |    |
Db        915 GGGNGCGCGGCGNGGCCCGCGCCCGCGCGCNCCCCCNCGCGCCGGCGNCGCGCGCCGNCC 974

Qy        684 CTCGGTGACCGGCGGCCG--CCTCGACACCCGCGAGCTAGACGCGGCGTACTGGTACCGC 741
              | |   | |   |||| |  || | | ||||    | || ||||  | |   |||| 
Db        975 CGCCCCGCGCCCGCGCCGGCCNCCGCCGCGCGCGCCGCGGCCGGGGCGGCCCGCGGCCGC 1034

Qy        742 AACATGTCGAGCACGGTCCGGTTCGAGCCCGCCGCCCGGCTGC 784
              |  | |   | ||    ||  || | || | |  |||| ||
Db       1035 CGCCCGGCCCCCCCGCCGGGGGCCGCGNCCNCGGGGCGGCCGC 1077
```

RESULT 9
AG032383/c
LOCUS AG032383 1169 bp DNA linear GSS 01-NOV-2001
DEFINITION Pan troglodytes DNA, clone: PTB-007B18.F, genomic survey sequence.
ACCESSION AG032383
VERSION AG032383.1 GI:16559256
KEYWORDS GSS.
SOURCE Pan troglodytes male lymphoblast DNA, clone_lib:PTB Chimpanzee Male BAC Library clone:PTB-007B18.F.
ORGANISM Pan troglodytes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE 1
AUTHORS Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE BAC end sequences of Library PTB
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1169)
AUTHORS Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE Direct Submission
JOURNAL Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC); 1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:chimpbes@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/, Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT Clones are derived from the chimpanzee BAC library PTB This BAC end was generated during the R&D process and may have higher chance of clone tracking errors.
PRIMERS
Sequencing: -21M13
LIBRARY
Vector : pKS145
R.Site 1 : SacI
R.Site 2 : SacI.

```
FEATURES             Location/Qualifiers
     source          1. .1169
                     /organism="Pan troglodytes"
                     /db_xref="taxon:9598"
                     /clone="PTB-007B18.F"
                     /sex="male"
                     /cell_type="lymphoblast"
                     /clone_lib="PTB Chimpanzee Male BAC Library"
BASE COUNT       32 a    244 c    627 g     25 t    241 others
ORIGIN
```

Query Match 8.2%; Score 78.2; DB 17; Length 1169;
Best Local Similarity 39.0%; Pred. No. 0.00011;
Matches 321; Conservative 0; Mismatches 497; Indels 5; Gaps 2;

```
Qy         74 TCCGGGAGAGCGTCCGCGCCTGCGAAGCCGCGTTCGCGCCCTACGTCGACTGGTCGGTGG 133
              | | || | |||  |  ||   |   | |    ||||| |     || |  | |||  |
```

```
Db       1138 TNCCGGCGCGCGNGCCNGCGCCCCCCCCGGNCGCCGCGCGCGGGCGCGCCGCGCCGGCCG 1079

Qy        134 AGCAGGTGTTGCGGGACTCGCCGGACGCTCCCGGGCTGGACCGGGTGG----ACGTCGTC 189
                || |   | |  |  | || ||| ||   || || ||| | ||    ||     |
Db       1078 CCGCGGCGCNCCCGCGCCGGNCGCGCCCCGCCGCCGGCCCGCGCGGCGCNCCGCGCCC 1019

Qy        190 CAGCCGACCCTGTTCGCCGTCATGATCTCCCTGGCCGCCCTCTGGCGCTCGCAAGGGGTC 249
              | || | |    |  |  |  | | || ||| |     || | |||  | || |
Db       1018 GCGGCGGCGCGCCGCCGCCCCGCGCCCGCNCTCGCCCGCGCGCCGCCCGCGCCCGCGGGC 959

Qy        250 GAGCCGTGCGCGGTGCTGGGACACAGCCTGGGCGAGATCGCGGCAGCCCACGTCTCGGGA 309
              ||||  ||| |  ||  |    | ||| | |      | |  |   || || | ||
Db        958 CCGCCGCGCGGGCCGCGCGCGNCCNGCCGGCGGCGCCGCCCCCCCCGCCCCGCCCCGCCC 899

Qy        310 GGCCTGTCCCTGGCCGACGCCGCACGCGTGGTGACGCTTTGGAGCCAGGCACAGACCACC 369
              ||||   ||| | |   || |   |||  |  | |    |  ||   | || | | ||
Db        898 GGCCGCCCCCCGCCGNCCGGCCGCCGCCCCGCCCCCCCGCGCCCGCCCGCCGCGCCGCCC 839

Qy        370 CTTGCCGGGACCGGCGCGCTCGTCTCCGTCGCCGCCA-CGCCGGATGAGCTCCTGCCCCG 428
              || | | |||    | | | |  ||||||  | ||||  | || || || | |
Db        838 GCGCGCGCGCCGGCCGCCGCGGCCGCGGCGGCCGCCGGCCCCGGGCGCGCGCCCGCGCGC 779

Qy        429 AATCGCTCCGTGGACCGAGGACAACCCGGCGCGGCTCGCCGTCGCAGCCGTCAACGGACC 488
              |||  || |  ||  |    || |||  ||  | |   | || |    ||
Db        778 GCGCGCGGCGCGCGGCGGCGCGGCGCCCGCGGCGCGGGGCCCGGGGGCGGCGCCCCCGCC 719

Qy        489 CCGGAGCACAGTCGTTTCCGGTGCCCGCGAGGCCGTCGCGGACCTGGTGGCCGACCTCAC 548
              | |       |||| | |  || ||||  ||| | | |  |    | |
Db        718 NGCGCGNCNCCCGCNCCCGGCGGCGCCGCCGCCGCGCCGGGNCNCGGGCGCCCNGCCGC 659

Qy        549 CGCCGCGCAGGTGCGCACGCGCATGATCCCGGTGGACGTTCCCGCCCACTCCCCCCTGAT 608
              | |||    |  | |   |    ||||      |  | |   |    ||  |
Db        658 CNCCGNCGCGCGCCCCGNCCNNNNNGCCCCGNNNCGNGNGCNCNNNCCNCGCCGNCCCCG 599

Qy        609 GTACGCCATCGAGGAACGGGTCGTCAGCGGCCTGCTGCCCATCACCCCACGCCCCTCCCG 668
                    |    |  |  |    |  | |||  | |||   ||||  |||
Db        598 NNNNCNNNNCNCCNCNCCNNCCCCCCNNCNNCCNCNCCCCCCCCCCCCNCNCCCCCCCCN 539

Qy        669 CATCCCCTTCCACTCCTCGGTGACCGGCGGCCGCCTCGACACCCGCGAGCTAGACGCGGC 728
              | | ||  |  | || |    ||  |  || ||          ||       |
Db        538 CNCCNCCNCCNNCCCCCCNNNNCCCCCCNCCCNCCCNCNNCNCCNNNGNNNNCCCNNNCN 479

Qy        729 GTACTGGTACCGCAACATGTCGAGCACGGTCCGGTTCGAGCCCGCCGCCCGGCTGCTTCT 788
              |    |          |  | |           || |  | | |  |  |  |
Db        478 NNCCCCNCNCNNNNCNCCCCNNNNCNNNNCCGGGCCNCNNCCNCCNNCNCNCCNNCCCNN 419

Qy        789 GCAGCAGGGGCCCAAGACGTTCGTCGAGATGAGCCCGCACCCGGTGCTGACCATGGGCCT 848
              |      ||||       |  |      |||| |  ||  ||  ||      |
Db        418 CCCCNNCCCGCCCCNCGNNNCCCCGGCCNNNCGCCCCCCNNCGCCGCNCGCCNCNCCGCC 359

Qy        849 CCAGGAGCTCGCCCCGGACCTGGGCGACACCACCGGCACCGCC 891
              |   ||  ||||  ||  || | |      |  | ||
Db        358 CGCCCCGCGNCCCCCCCCCCCCCCCGCCCCNCNNNCCNNCCCC 316
```

RESULT 10
AG060835/c
LOCUS AG060835 1435 bp DNA linear GSS 03-NOV-2001
DEFINITION Pan troglodytes DNA, clone: PTB-048L05.R, genomic survey sequence.
ACCESSION AG060835
VERSION AG060835.1 GI:16612637
KEYWORDS GSS.
SOURCE Pan troglodytes male lymphoblast DNA, clone_lib:PTB Chimpanzee Male BAC Library clone:PTB-048L05.R.
ORGANISM Pan troglodytes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE 1
AUTHORS Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE BAC end sequences of Library PTB
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1435)

```
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            (E-mail:chimpbes@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT     Clones are derived from the chimpanzee BAC library PTB This BAC end
            was generated during the R&D process and may have higher chance of
            clone tracking errors.
             PRIMERS
                 Sequencing: M13Rev
             LIBRARY
                 Vector    : pKS145
                 R.Site 1  : SacI
                 R.Site 2  : SacI.
FEATURES             Location/Qualifiers
     source          1. .1435
                     /organism="Pan troglodytes"
                     /db_xref="taxon:9598"
                     /clone="PTB-048L05.R"
                     /sex="male"
                     /cell_type="lymphoblast"
                     /clone_lib="PTB Chimpanzee Male BAC Library"
BASE COUNT       97 a    551 c    565 g     17 t    205 others
ORIGIN

  Query Match             8.2%;  Score 78.2;  DB 17;  Length 1435;
  Best Local Similarity   43.5%;  Pred. No. 0.00011;
  Matches  318;  Conservative    0;  Mismatches  405;  Indels    8;  Gaps    2;

Qy         55 CTCGACGCTTCCGACGTCTTCCGGGAGAGCGTCCGCGCCTGCGAAGCCGCGTTCGCGCCC 114
              |  | |||   |  ||    ||||| | ||  |  | |  |||  ||||    | | ||
Db       1164 CGNGCCGCCNNCCGCGCNCGCCGGGCGCGCCCCGCCCCGNGCGCCGCCG----CNCNCCN 1109

Qy        115 TACGTCGACTGGTCGGTGGAGCAGGTGTTGCGGGACTCGCCGGACGCTCCCGGGCTGGAC 174
              || || |    |  || ||  |    ||||  | |||   | ||| ||    | | |
Db       1108 GCCGGCGNCGNCGCNCCGGCGCGCGCNCCGCGGCTCGCGNNCGCCGCGCCNNCCCCGCGC 1049

Qy        175 CGGGTGGACGTCGTCCAGCCGACCCTGTTCGCCGTCATGATCTCCCTGGCCGCCCTCTGG 234
              |  | || ||     || | |    |       |||  | |  | |
Db       1048 GCCGCCGGCGCCGGGGCCGCGNCGCCCNNNGNGCGGCCCGGCTCNGCGTCGNCNCNNNCC 989

Qy        235 CGCTCGCAAGGGGTCGAGCCGTGCGCGGTGCTGGGACACAGCCTGGGCGAGATCGCGGCA 294
              || |    | |    |  ||||  | ||  |  | | |||   | | |  || |||
Db        988 GGCGCCGCCGCCGCGCGGGGCCGCGCNGGGCGCGCGCGCNGCCGNCGTGTGGTCCCGGTG 929

Qy        295 GCCCACGTCTCGGGAGGCCTGTCCCTGGCCGA----CGCCGCACGCGTGGTGACGCTTTG 350
              |  | || | || | | |    | | | ||      || ||   ||| || | |    |
Db        928 GGNCNCGGCGCGCGCGTCTCCGCNCGCGTCGCGGGCCGGCGGCGGCGCGGCGGGGGCGGG 869

Qy        351 GAGCCAGGCACAGACCACCCTTGCCGGGACCGGCGCGCTCGTCTCCGTCGCCGCCACGCC 410
              | |  ||    |  ||  | || |  |||| |||    | |||  || |||| |||
Db        868 CGGGCGCGCCNCNNCGNCCGGCGGCGCGCGCGGCNCGCGNCGCGCCGGCGGCGCCGCGCG 809

Qy        411 GGATGAGCTCCTGCCCCGAATCGCTCCGTGGACCGAGGACAACCCGGCGCGGCTCGCCGT 470
              |  ||  |    | |      | |   |   || |       ||  |||  |   ||
Db        808 CCGGGCGCGGCCCGGCGGCGCGNCGCGTGGCGGCGCGCCGCCGGCGCGGCGCCCNCCCCG 749

Qy        471 CGCAGCCGTCAACGGACCCCGGAGCACAGTCGTTTCCGGTGCCCGCGAGGCCGTCGCGGA 530
              |||  |       ||| | |||  || | |  |    |      |||||  |  ||| |
Db        748 CGCCGNGCCGCGCGGGCGCCGCCGCGCCGCGGCGCCTCTCCGCCGCGCGCGCGCCGCCCC 689

Qy        531 CCTGGTGGCCGACCTCACCGCCGCGCAGGTGCGCACGCGCATGATCCCGGTGGACGTTCC 590
              |  ||  ||    | |  ||||||  |   ||  ||    |  |||   ||      |
Db        688 GCCGGGCGCGCGGGGCGCGCCCGCGCGCGCCGGCGGGCCGCGGCGCCCCCGGGCGCGNCG 629

Qy        591 CGCCCACTCCCCCCTGATGTACGCCATCGAGGAACGGGTCGTCAGCGGCCTGCTGCCCAT 650
              |  ||  || || |       ||    || |   ||   |    | |||  |   ||
Db        628 CNNCCGGTCGCCGCGCGGCGGCGGGGCCGGGCGGCGCCGCCNGGGGCGCCGNCNCGCCGC 569

Qy        651 CACCCCACGCCCCTCCCGCATCCCCTTCCACTCCTCGGTGACCGGCGGCCGCCTCGACAC 710
```

```
              |||   | ||   ||||  |||  || | || ||      ||   |||  | |  | |
Db        568 GCCCCGCGGGCCGCGCCGCGCNCCCCCCCCCGCCGCGCGCCGCGCGCCCGCCCCGCCC 509

Qy        711 CCGCGAGCTAGACGCGGCGTACTGGTACCGCAACATGTCGAGCACGGTCCGGTTCGAGCC 770
              ||||     |  |||  |    ||   | |  |  | |  || ||  |  |  ||  ||
Db        508 GCGCGGCGGCGCGGCGCGGCNNGGGGGGCCCGGCGGGCCCCGCCCGCCCGCGGGCGCCCC 449

Qy        771 CGCCGCCCGGC 781
              |||  |||  |
Db        448 CGCGCCCCCNC 438


RESULT 11
AG058791
LOCUS       AG058791                842 bp    DNA     linear   GSS 02-NOV-2001
DEFINITION  Pan troglodytes DNA, clone: PTB-045M15.F, genomic survey sequence.
ACCESSION   AG058791
VERSION     AG058791.1  GI:16596252
KEYWORDS    GSS.
SOURCE      Pan troglodytes male lymphoblast DNA, clone_lib:PTB Chimpanzee Male
            BAC Library clone:PTB-045M15.F.
  ORGANISM  Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE   1
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     BAC end sequences of Library PTB
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 842)
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            (E-mail:chimpbes@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT     Clones are derived from the chimpanzee BAC library PTB This BAC end
            was generated during the R&D process and may have higher chance of
            clone tracking errors.
             PRIMERS
                 Sequencing: -21M13
             LIBRARY
                 Vector    : pKS145
                 R.Site 1  : SacI
                 R.Site 2  : SacI.
FEATURES             Location/Qualifiers
     source          1. .842
                     /organism="Pan troglodytes"
                     /db_xref="taxon:9598"
                     /clone="PTB-045M15.F"
                     /sex="male"
                     /cell_type="lymphoblast"
                     /clone_lib="PTB Chimpanzee Male BAC Library"
BASE COUNT       36 a    396 c    340 g     26 t     44 others
ORIGIN

  Query Match             8.1%;  Score 76.4;  DB 17;  Length 842;
  Best Local Similarity   46.4%;  Pred. No. 0.00022;
  Matches  324;  Conservative    0;  Mismatches  361;  Indels   13;  Gaps    3;

Qy         32 GGCCGGGCATGGGAAGGGAACTTCTCGACGCTTCCGACGTCTTCCGGGAGAGCGTCCGCG 91
              || ||||   | |  |||  |       | ||       | |  || |  | ||  |  ||
Db         88 GGGCGGGGNNGNGNNGGGGGCGNGCTTTCTCTCGGAGGGGCGCCCCGCGGCGCCGCGCCG 147

Qy         92 CCTGCGAAGCCGCGTTCGCGCCCTACGTCGACTGGTCGGTGGAGCAGGTGTTGCGGGACT 151
                 |    ||   | ||||  |    |  | ||  | || || |  |||||||
Db        148 GGGCCCCCCGCGGCCGGGGGCCCGCCCGGCCCCCGCCGCCCGCGCCGGGGGCGCGGGACC 207

Qy        152 CGCCGGACGCTCCCGGGCTGGACCGGGTGGACGTCGTCCAGCCGACCCTGTTCGCCGTCA 211
                |   | ||| |||  ||   || ||  | ||       ||     |  |||
Db        208 GGGGTTTTGTTCCGGGGGCGGCGGCGGCGGCGGGCGAGGGAAAGAAAAAAAGCAACGTTG 267
```

```
Qy        212 TGATCTCCCT-GGCCGCCCTCTGGCGCTCGCAAGGGGTCGAGCCGTGCGCGGTGCTGGGA 270
              | | |  |||||  | | | |  |     |  |||| ||||  |  ||
Db        268 CCTGCCCGTTCCNCCGCCGGCGGCCCCGGGGGGCCGCCCCGGCCGGCCGCGCGGGGGGNG 327

Qy        271 CACAGCCTGGGCGAGATCGCGGCAGCCCACGTCTCGGGAGGCCTGTCCCTGGCCGACGCC 330
              |||  | |||||     |||||  | | |  | || | ||   | |||     || | ||
Db        328 NACACGCCGGGCGCCGCGGCGGCGCCGCGCCGCCCGCGCGGGGCGGCCC---GCGCCCCC 384

Qy        331 GCACGCGTGGTGACGCTTTGGAGCCAGGCACAGACCACCCTTGCCGGGACCGGCGCGCTC 390
              || |||| |  | | |    | ||| |   |   || |||  | | || ||  | ||| |
Db        385 GCTCGCGCGCCGGCCCCGGCGGGCCCGCGCCCCGCCCCCCCCGGCCGGCCCCCCCCGCGC 444

Qy        391 GTCTCCGTCGCCGCCACGCCGGATGAGCTCCTGCCCCGAATCGCTCCGTGGACCGAGGAC 450
              |   ||   |||| | | ||          | ||||||   ||| ||     | | ||
Db        445 GGNNCCCCGGCCGGCGCCCC---------CGCGCCCCGCCCCGCGCCCGCCGCGGCGGCG 495

Qy        451 AACCCGGCGCGGCTCGCCGTCGCAGCCGTCAACGGACCCCGGAGCACAGTCGTTTCCGGT 510
              | | ||  |||  |||| ||  ||||     |||  |||     |   |     |||
Db        496 CGCGCCGCCGGGCCGGCCGCCGGCGCCGCGCCCGGCGCCCCCGCCCGGGCGCGGCCCGCG 555

Qy        511 GCCCGCGAGGCCGTCGCGGACCTGGTGGCCGACCTCACCGCCGCGCAGGTGCGCACGCGC 570
              | |||   |||| ||  |  |  |    |||| || |||||  |   | |   | | ||| |
Db        556 NCGCGCCGCGCCGGCGGGCGCGCGCGCGCCGCCCCCACCGGCCGCCCGCCCCCCCCGCCC 615

Qy        571 ATGATCCCGGTGGACGTTCCCGCCCACTCCCCCCTGATGTACGCCATCGAGGAACGGGTC 630
              |  | ||| || ||  || |  | | ||| ||       | ||  || ||      |
Db        616 CGGGCCGCGGNGGGCGCCCCGGNGCCCGCCCGCCCCGGANNCCCCCCCGCGGCGGCCGGG 675

Qy        631 GTCAGCGGCCTGCTGCCCATCACCCCACGCCCCTCCCGCATCCCCTTCCACTCCTCGGTG 690
              | ||| |  |  |   |  | |   ||  | ||  |||    |  ||  |  |
Db        676 CGCGGCGCCGCGGGCCGGCCCGGCGCGGNNCCGCGCGGCGCCCCGGCGCGGCCCCGGCCG 735

Qy        691 ACCGGCGGCCGCCTCGACACCCGCGAGCTAGACGCGGC 728
              |  ||  |||||| | ||  ||  ||  |    || ||
Db        736 AGGGGNCGCCGCCCCCACGCCGCCGGNCCGCGAGCCGC 773


RESULT 12
CNS010RY
LOCUS       CNS010RY                776 bp    DNA     linear   GSS 26-JUL-1999
DEFINITION  Drosophila melanogaster genome survey sequence SP6 end of BAC
            BACN04P24 of DrosBAC library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL099352
VERSION     AL099352.1  GI:5610963
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 776)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the European Drosophila Genome Project (EDGP) -
            http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
            library (Dros BAC) was made by Alain Billaud at CEPH (Centre
            d'Etude du Polymorphisme Humain) with funding provided by a MRC
            project grant. The DNA was prepared from embryos by Alain Bucheton
            and Genevieve Payan. It has been constructed in the vector
            pBeloBAC11.
FEATURES             Location/Qualifiers
     source          1. .776
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone="BACN04P24"
                     /clone_lib="DrosBAC"
                     /plasmid="pBeloBAC11"
                     /note="end : SP6"
BASE COUNT      110 a    439 c    135 g      0 t     92 others
ORIGIN

  Query Match             8.0%;  Score 76;  DB 17;  Length 776;
  Best Local Similarity   39.1%;  Pred. No. 0.00026;
  Matches  190;  Conservative   53;  Mismatches  243;  Indels    0;  Gaps    0;

Qy        254 CGTGCGCGGTGCTGGGACACAGCCTGGGCGAGATCGCGGCAGCCCACGTCTCGGGAGGCC 313
              :|  ||||     |   |  |  ||   ||: |  |:|  |  ||| :: | |:     ||
Db         12 MGCCCGCGCGCAGACSCCCGCCCCCCGCSCGCCCSCCCCSCCCCGSSCCCCSCCCCCCC 71

Qy        314 TGTCCCTGGCCGACGCCGCACGCGTGGTGACGCTTTGGAGCCAGGCACAGACCACCCTTG 373
                :::  : |: |  ::| ||| :    ::|   :   || | ||  | || |:|  |
Db         72 CCGSSSCCSGCSCCCGSSCGCGCCCSCSCGSSCGSCSCGCCCCGCCAGCGCCCSCSCGGG 131

Qy        374 CCGGGACCGGCGCGCTCGTCTCCGTCGCCGCCACGCCGGATGAGCTCCTGCCCCGAATCG 433
              ||:   ||    ||| || | ||  ::|| || |:|||       |  :  | ||    |
Db        132 CCSMMSCCCCGCCGCGCGCCCCCCCSSCCCCCCCSCCGCCCCCCCGGGSSCCGCCCSCCCC 191

Qy        434 CTCCGTGGACCGAGGACAACCCGGCGCGGCTCGCCGTCGCAGCCGTCAACGGACCCCGGA 493
              | ||   || ||  | || ||||  ::|||  || | | | :||| |  : : |:|:
Db        192 CCCCCCCGAGCGCCGGCACCCCGCGSMGGCCGGCAGSCCCCSCCGSCGSSCSSCSCSCCC 251

Qy        494 GCACAGTCGTTTCCGGTGCCCGCGAGGCCGTCGCGGACCTGGTGGCCGACCTCACCGCCG 553
              :| | |       :|    ||: |    ||  | |   || :    :||  | | ||:||:
Db        252 SCCCCGCGCSCGSCCCGCCCSCCCCCCCCCCGCCCCCSCCCSCCCCSCGSGCCCCCCSCCS 311

Qy        554 CGCAGGTGCGCACGCGCATGATCCCGGTGGACGTTCCCGCCCACTCCCCCCTGATGTACG 613
              | | :   ||| | ||:  :   |  |    |: ||| ||  | ||| ::       |
Db        312 CCCCSCCCCGCSCCCGSCGSSCGCGCGCCCCCSCCCCCCCCGCCCCCCGSSSCSCCCCCC 371

Qy        614 CCATCGAGGAACGGGTCGTCAGCGGCCTGCTGCCCATCACCCCACGCCCCTCCCGCATCC 673
              |  |     || | || : :|      |  :||    |||  ::|||| |||:|  ||
Db        372 GCCCCCCCCCCCCGCGSCGCSSSCCCGGSCCCCSCCCCACCCCAGSSCCCCGCCCSCCCCC 431

Qy        674 CCTTCCACTCCTCGGTGACCGGCGGCCGCCTCGACACCCGCGAGCTAGACGCGGCGTACT 733
              |  || : :| |::   | |  ||||  | |  |||||||    |     | |  || ||
Db        432 GCCGCCCSCSCCCSSCCCCGGCGGGCCCGCGCCCCACCCGCACCCGCACCACCCCGCACC 491

Qy        734 GGTACC 739
              |  ||
Db        492 CGCCCC 497


RESULT 13
CNS015Y4/c
LOCUS       CNS015Y4                1203 bp    DNA     linear   GSS 26-JUL-1999
DEFINITION  Drosophila melanogaster genome survey sequence T7 end of BAC
            BACN15E10 of DrosBAC library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL106054
VERSION     AL106054.1  GI:5619805
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1203)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the European Drosophila Genome Project (EDGP) -
            http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
            library (Dros BAC) was made by Alain Billaud at CEPH (Centre
            d'Etude du Polymorphisme Humain) with funding provided by a MRC
            project grant. The DNA was prepared from embryos by Alain Bucheton
            and Genevieve Payan. It has been constructed in the vector
            pBeloBAC11.
```

```
FEATURES             Location/Qualifiers
     source          1. .1203
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone="BACN15E10"
                     /clone_lib="DrosBAC"
                     /plasmid="pBeloBAC11"
                     /note="end : T7"
BASE COUNT      154 a    274 c    380 g    158 t    237 others
ORIGIN

  Query Match             8.0%;  Score 76;  DB 17;  Length 1203;
  Best Local Similarity   34.0%;  Pred. No. 0.00026;
  Matches  177;  Conservative   90;  Mismatches  254;  Indels    0;  Gaps    0;

Qy        139 GTGTTGCGGGACTCGCCGGACGCTCCCGGGCTGGACCGGGTGGACGTCGTCCAGCCGACC 198
              | |  | |||    ||:||   | :| :|:|  : : |:  :  :| :|    || | |:
Db       1196 GGGGGGGGGGGGGGGCSGGSGCCCSCGSGSCSCSGSGGSCCSCGSGGSGGGGGCGGCCS 1137

Qy        199 CTGTTCGCCGTCATGATCTCCCTGGCCGCCCTCTGGCGCTCGCAAGGGGTCGAGCCGTGC 258
              |  | |||    |  | ||| ::: |:   | || |:  ||  | ::  | |:|  |
Db       1136 GGGGGCCCCGGGCGGCCCCCCCSSSGGSGGSCSGGGGSGGGCCGGCSSSGGGGSCCGGG 1077

Qy        259 GCGGTGCTGGGACACAGCCTGGGCGAGATCGCGGCAGCCCACGTCTCGGGAGGCCTGTCC 318
              |:|| ::  :: : : |      |     | | :: :||| :: | |  : ::|::  ||
Db       1076 GSGGGSSCCSSSSSSGGGGCCCCCCCCCCCCCCSSSSCCCSSKCCCCCCKNSKCSKCCCC 1017

Qy        319 CTGGCCGACGCCGCACGCGTGGTGACGCTTTGGAGCCAGGCACAGACCACCCTTGCCGGG 378
              |  :: | | || | |::   :  | |       |: ::| : : :: :||: |:
Db       1016 CSCSSGGGCCCCCCCCKBCGCCBCCCCCCSSCCCCCSCSSCCSBSSSSBBCCKGGSKCCC 957

Qy        379 ACCGGCGCGCTCGTCTCCGTCGCCGCCACGCCGGATGAGCTCCTGCCCCGAATCGCTCCG 438
              |: :| ::| |: | |:::| |:::| |:||::    | |   ||||   | | :|
Db        956 CCBTKCCBKCCCKSCCCBKBCCCBBBCCCSCCSBKCCCCCCCNCCCCCCCCCCCCCCSCC 897

Qy        439 TGGACCGAGGACAACCCGGCGCGGCTCGCCGTCGCAGCCGTCAACGGACCCCGGAGCACA 498
                : :   :: |  |||:| |:||  | ||  | |          || |||:||  | |
Db        896 CCSSSGCGSSCCCCCCCSGGGSGGGNCCCCCCCCCCCGGCSGGGGGGSCCCSGGGCCCCC 837

Qy        499 GTCGTTTCCGGTGCCCGCGAGGCCGTCGCGGACCTGGTGGCCGACCTCACCGCCGCGCAG 558
              |:    |    || | | | ||| : |   || ||   :|  || | ||:|:| | |
Db        836 CCCSGSGGCCCSCCCGGGGGGCCCGCSCCCCCCCSGGCCCSCCCCCCCSCCSCSGGGGGG 777

Qy        559 GTGCGCACGCGCATGATCCCGGTGGACGTTCCCGCCCACTCCCCCCTGATGTACGCCATC 618
              | | |  | |  |    || :| ||    | |:| | |||||:        | ||  |
Db        776 GGGGGGGGGGGGGGGGGGGGGGRGCCGSCGGCCCSCCCCCCCCCSCCCCCCCCCCCSC 717

Qy        619 GAGGAACGGGTCGTCAGCGGCCTGCTGCCCATCACCCCACG 659
              : ||  |    |:   | ||: |: |:::  :  :|| ||
Db        716 SCGGCGCCCCCCSSGGGGGGSGGGSSGSSSGSSSGSCCSCG 676


RESULT 14
BG844318/c
LOCUS       BG844318                873 bp    mRNA    linear   EST 29-MAY-2001
DEFINITION  1024005H12.x2 C. reinhardtii CC-1690, normalized, Lambda Zap II
            Chlamydomonas reinhardtii cDNA, mRNA sequence.
ACCESSION   BG844318
VERSION     BG844318.1  GI:14225502
KEYWORDS    EST.
SOURCE      Chlamydomonas reinhardtii.
  ORGANISM  Chlamydomonas reinhardtii
            Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
            Chlamydomonadaceae; Chlamydomonas.
REFERENCE   1  (bases 1 to 873)
  AUTHORS   Grossman,A., Davies,J., Federspiel,N., Harris,E., Lefebvre,P.,
            McDermott,J.P., Silflow,C., Stern,D. and Surzycki,R.
  TITLE     Analyses of the Chlamydomonas reinhardtii Genome: A Model,
            Unicellular System for Analyzing Gene Function and Regulation in
            Vascular Plants; project phase 2
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Charles Hauser
            DCMB Box 91000
            Duke University
            Durham, NC 27708-1000
            Tel: 919 613 8159
            Fax: 919 613 8177
            Email: chauser@duke.edu.
FEATURES             Location/Qualifiers
     source          1. .873
                     /organism="Chlamydomonas reinhardtii"
                     /strain="CC-1690 wild type mt+ 21gr"
                     /db_xref="taxon:3055"
                     /clone_lib="C. reinhardtii CC-1690, normalized, Lambda Zap
                     II"
                     /note="Vector: pBluescript II SK-; Site_1: EcoRI; Site_2:
                     XhoI; This library, constructed by John Davies and Jeffrey
                     McDermott, combines cDNAs from CC-1690 cells grown to
                     mid-log phase in TAP (acetate-containing) medium in the
                     light, TAP medium in the dark, HS (minimal) medium in
                     ambient levels of CO2 and HS medium bubbled with 5% CO2.
                     PolyA mRNA was purified from each sample, pooled and cDNA
                     synthesized. The cDNA was directionally cloned into lambda
                     ZAP II (Stratagene) in the EcoRI (5') and XhoI (3') sites.
                     pBluescript II SK- plasmids were excised from the lambda
                     ZAP clones by superinfection with ExAssist (Stratagene)
                     phage. The library was normalized using method 4 described
                     in Bonaldo et al (1996) Genome Research 6: 791-806."
BASE COUNT       17 a    334 c    471 g     22 t     29 others
ORIGIN

  Query Match             8.0%;  Score 75.4;  DB 12;  Length 873;
  Best Local Similarity   45.8%;  Pred. No. 0.00034;
  Matches  286;  Conservative    0;  Mismatches  336;  Indels    2;  Gaps    1;

Qy         87 CCGCGCCTGCGAAGCCGCGTTCGCGCCCTACGTCGACTGGTCGGTGGAGCAGGTGTTGCG 146
              ||||||   ||  ||| ||      | |  || |    |  | |  | |      |  ||
Db        742 CCGCGCGCNCGCGGCCCCGCCGCGCCGCCCCGCCCCGCGCCCCGCCGGGGGCCGGCGGCC 683

Qy        147 GGACTCGCCGGACGCTCCCGGGCTGGACCGGGTGGACGTCGTCCAGCCGACCCTGTTCGC 206
              ||   |||||| | | |||| ||    || |  || ||     |  |||| ||| |  |||
Db        682 GGNGGCGCCGGCCCCCCCCGCGCCCCCCCCGCCGGGCGCGCCCGCGCCGCCCCGGCCCGC 623

Qy        207 CGTCATGATCTCCCTGGCCGCCCTCTGGCGCTCGCAAGGGGTCGAGCCGTGCGCGGTGCT 266
              |  |  |  |||| | |  | | | |    |  |||  || | ||
Db        622 GGCGCGCGGGGCGCGCGCCGGGCGCCCCCCCGCCCCCCCGCCGCCCCCGCCGGCCGGGCG 563

Qy        267 GGGACACAGCCTGGGCGAGATCGCGGCAGCCCACGTCTCGGGAGGCCTGTCCCTGGCCGA 326
              || | | ||  ||  | |  |||| | |||  |  |    |     | |  |  ||||
Db        562 CGGCCGCCGCGCGGCGGCGGCCGCGCCGGCCGCCCCCCGCGCCCCGCGGCGCGCGGCCCG 503

Qy        327 CGCCGCACGCGTGGTGACGCTTTGGAGCCAGGCACAGACCACCCTTGCCGGGACCGGCGC 386
              ||| || | || || | |||    |    || | | |  |||  | | | | |||
Db        502 CGCGGCCCCCGCGGCGGCGCCCCGCGCCGGCGCGCCGCCGCCCCCGCGGGCGCCGGGCGC 443

Qy        387 GCTCGTCTCCGTCGCCGCCACGCCGGATGAGCTCCTGCCCCGAATCGCTCCGTGGACCGA 446
              |  | |  | || || | | ||    ||    ||| ||   | | | | | ||
Db        442 CGCCCCCCCGCCCCCCCCCGCCCGGGCCCCGCGGGCGCCGCGCCCCCCCCGGGGCGCCCG 383

Qy        447 GGACAACCCGGCGCGGCTCGCCGTCGCAGCCGTCAACGGACCCCGGAGCACAGTCGTTTC 506
              |       ||||||   |||  | |  ||  |  ||      ||    | |  |
Db        382 GCGGGGGGGGGCGCGCGGGGCCCGCCCCCCCCGCGCGGGGGGGGGGGCGGCGGCGGGGCG 323

Qy        507 CGGTGCCCGCGAGGCCGTCGCGGACCTGGTGGCCGACCTCACCGCCGCGCGCAGGTGCGCAC 566
              |  | |  ||  | | | || ||  |                 ||||  |  |  | | |
Db        322 GCCGGGCGGGGCCGCGCGCCCCGCCCGGGCCGGGCCGGCGCGGGCCGCCCGCGCCCCCGC 263

Qy        567 GCGCATGATCCCGGTGGACGTTCCCGCCCACTCCCCCCTGATGTACGCCATCGAGGAACG 626
              ||||  |  |||||  |     ||||||  | |||  |      |||  |    |  |
Db        262 GCGCGCGCGCCCGGGCGCAACGCCCGCC--CCCCCNGCAANNTCACGTAAATTTTGCGGG 205

Qy        627 GGTCGTCAGCGGCCTGCTGCCCATCACCCCACGCCCCTCCCGCATCCCCTTCCACTCCTC 686
              ||   | || ||   |  ||     |||| ||||||  || |  ||||   | | ||
Db        204 GGGGCTGGGCCGCGCCCGCCCGGCGGCCCCCCGCCCCCGCCCCCCCCCCGGGCGCGCCCG 145

Qy        687 GGTGACCGGCGGCCGCCTCGACAC 710
```

```
Db        144 GGGCGCCCCCGCCCCGCCCCCCC 121

RESULT 15
CNS077R7
LOCUS       CNS077R7                753 bp    DNA     linear   GSS 08-JUL-2001
DEFINITION  T7 end of clone BB0AA014A04 of library BB0AA from strain CBS 4732
            of Pichia angusta, genomic survey sequence.
ACCESSION   AL433049
VERSION     AL433049.1  GI:12216463
KEYWORDS    GSS.
SOURCE      Pichia angusta.
  ORGANISM  Pichia angusta
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; Saccharomycetaceae; Pichia.
REFERENCE   1  (bases 1 to 753)
  AUTHORS   Souciet,J.L., Aigle,M., Artiguenave,F., Blandin,G.,
            Bolotin-Fukuhara,M., Bon,E., Brottier,P., Casaregola,S.,
            de-Montigny,J., Dujon,B., Durrens,P., Lepingle,A., Llorente,B.,
            Malpertuy,A., Neuveglise,C., Ozier-Kalogeropoulos,O., Potier,S.,
            Saurin,W., Tekaia,F., Toffano-Nioche,C., Wesolowski-Louvel,M.,
            Wincker,P. and Weissenbach,J.
  TITLE     Genomic exploration of the hemiascomycetous yeasts: 1. A set of
            yeast species for molecular evolution studies
  JOURNAL   FEBS Lett. 487 (1), 3-12 (2000)
  MEDLINE   20584711
   PUBMED   11152876
REFERENCE   2  (bases 1 to 753)
  AUTHORS   Blandin,G., Llorente,B., Malpertuy,A., Wincker,P., Artiguenave,F.
            and Dujon,B.
  TITLE     Genomic exploration of the hemiascomycetous yeasts: 13. Pichia
            angusta
  JOURNAL   FEBS Lett. 487 (1), 76-81 (2000)
  MEDLINE   20584723
   PUBMED   11152888
REFERENCE   3  (bases 1 to 753)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-SEP-2000) Genoscope - Centre National de Sequencage,
            2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :
            seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT     This GSS is part of a random genomic sequencing program of thirteen
            yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces
            exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii,
            Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces
            lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia
            angusta, Debaryomyces hansenii var.  hansenii, Pichia sorbitophila,
            Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to
            5 kb were prepared and both extremities were sequenced. See
            keywords for description of this sequence and for the sequence of
            the other extremity of this insert.
FEATURES             Location/Qualifiers
     source          1. .753
                     /organism="Pichia angusta"
                     /strain="CBS 4732"
                     /db_xref="taxon:4905"
                     /clone="BB0AA014A04"
                     /clone_lib="BB0AA"
                     /note="end : T7"
     misc_feature    <604. .>753
                     /note="similar to Saccharomyces cerevisiae ORF YPR074c [
                     TKL1 ; transketolase 1 ]"
                     /evidence=not_experimental
BASE COUNT       40 a    464 c    123 g     82 t     44 others
ORIGIN

  Query Match           7.9%;  Score 75;  DB 17;  Length 753;
  Best Local Similarity   45.3%;  Pred. No. 0.00039;
  Matches  189;  Conservative   19;  Mismatches  209;  Indels    0;  Gaps    0;

Qy        374 CCGGGACCGGCGCGCTCGTCTCCGTCGCCGCCACGCCGGATGAGCTCCTGCCCCGAATCG 433

Db         92 CCCGGCCGGCCCCYYMCCCCCCGCCCCCCMYCCCCCCSCCCCCCCSTCCMMCCCCCCCTCC 151

Qy        434 CTCCGTGGACCGAGGACAACCCGGCGCGGCTCGCCGTCGCAGCCGTCAACGGACCCCGGA 493

Db        152 CCCCCGGCCGCCYCTTSSCCSCCCCSCGCGCCCCCGCCCCGGCGCCCCCCGCGCCCCCCC 211

Qy        494 GCACAGTCGTTTCCGGTGCCCGCGAGGCCGTCGCGGACCTGGTGGCCGACCTCACCGCCG 553

Db        212 CCYCCCTCCCCSCCCCCCCCCGCSCGGCTGCCGTGCSCCCGGCCCYCCGCGCCCCGCCCG 271

Qy        554 CGCAGGTGCGCACGCGCATGATCCCGGTGGACGTTCCCGCCCACTCCCCCCTGATGTACG 613

Db        272 CCTCCGCGCCCCCCCCCCTSCCCCTCCCCCCCCCCSCCTCCCCCGCCCGCCCGTCYTYCC 331

Qy        614 CCATCGAGGAACGGGTCGTCAGCGGCCTGCTGCCCATCACCCCACGCCCCTCCCGCATCC 673

Db        332 SCAGCCCCGCCCCCCYCCSTCCCCCCCCCGCCCCTGCCCCCCSCCCCCCCCCCCCCCGCC 391

Qy        674 CCTTCCACTCCTCGGTGACCGGCGGCCGCCTCGACACCCGCGAGCTAGACGCGGCGTACT 733

Db        392 CCCCCCCCCCCCGCGCSCCCCSSCCSCCGCCCCCGCCGCCCCGCGCCCCCCCCGGGCCCCC 451

Qy        734 GGTACCGCAACATGTCGAGCACGGTCCGGTTCGAGCCCGCCGCCCGGCTGCTTCTGC 790

Db        452 CGGCCGCCCCACCCCGGCGCCCGGCCCCCGCCSCCCCCGCCGCCCCCCCCCCSCCSC 508

Search completed: June 18, 2003, 01:00:03
Job time : 1221.33 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 17, 2003, 11:26:44 ; Search time 849.632 Seconds
(without alignments)
15003.005 Million cell updates/sec

Title: US-09-914-286-1_COPY_8947_9384
Perfect score: 438
Sequence: 1 tactggctcgaaagcacaca.....gtcatgcctcgggggtgctg 438

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		438	100.0	12381	6	AX006889	AX006889 Sequence
2		438	100.0	30690	6	E38020	E38020 Avermectin
3		438	100.0	64957	1	AB032367	AB032367 Streptomy
4		434.8	99.3	11096	1	AF275943	AF275943 Streptomy
5		431.6	98.5	438	6	E38022	E38022 Avermectin
6		273.6	62.5	31422	6	E38021	E38021 Avermectin
7	c	273.6	62.5	64957	1	AB032367	AB032367 Streptomy
8		154	35.2	113193	1	AF357202	AF357202 Streptomy
9		148.8	34.0	78210	1	AB070949	AB070949 Streptomy
10	c	139	31.7	104326	1	AB070940	AB070940 Streptomy
11	c	136	31.1	84985	1	SNA278573	AJ278573 Streptomy
12		133.6	30.5	37948	1	AF079138	AF079138 Streptomy
13		132.8	30.3	65140	6	AX211705	AX211705 Sequence
14		132.8	30.3	123580	1	AF263912	AF263912 Streptomy
15		132.8	30.3	125401	6	AX211739	AX211739 Sequence
16		131.2	30.0	27541	6	AX211706	AX211706 Sequence
17		130.4	29.8	77534	1	AF235504	AF235504 Streptomy
18	c	129.8	29.6	24225	1	SC2C4	AL512902 Streptomy
19		129.4	29.5	20394	1	SNA132222	AJ132222 Streptomy
20		129.4	29.5	20394	6	AX067996	AX067996 Sequence
21		126.2	28.8	104326	1	AB070940	AB070940 Streptomy
22	c	125.8	28.7	39314	1	SGR300302	AJ300302 Streptomy
23	c	124	28.3	1590	6	AR173222	AR173222 Sequence
24		123.4	28.2	16767	6	AX089464	AX089464 Sequence
25		123.4	28.2	50000	6	AX089417	AX089417 Sequence
26		123.4	28.2	50000	6	AX089420	AX089420 Sequence
27		123.4	28.2	80161	1	AY007564	AY007564 Saccharop
28		123.4	28.2	80161	6	AR165018	AR165018 Sequence
29		122.2	27.9	43280	1	SFU78289	U78289 Streptomyce
30		118.4	27.0	29736	6	AX089421	AX089421 Sequence
31		117.4	26.8	24568	1	AF220951	AF220951 Streptomy
32		117.4	26.8	50937	6	AR159871	AR159871 Sequence
33	c	117	26.7	107379	1	SHGCPIR	X86780 S.hygroscop
34	c	116.2	26.5	26195	1	SC1G7	AL591083 Streptomy
35		116	26.5	19791	1	SPFKBAD	Y10438 Streptomyce
36		116	26.5	30000	6	AX250262	AX250262 Sequence
37	c	114.4	26.1	77534	1	AF235504	AF235504 Streptomy
38	c	112.6	25.7	30000	6	AX250263	AX250263 Sequence
39		109.4	25.0	33529	6	AR166425	AR166425 Sequence
40		107.4	24.5	344050	1	MLEPRTN1	AL583917 Mycobacte
41		107	24.4	5676	6	A69718	A69718 Sequence 1
42		107	24.4	53784	1	AMM223012	AJ223012 Amycolato
43		107	24.4	53789	6	A69720	A69720 Sequence 3
44		107	24.4	90445	1	AF040570	AF040570 Amycolato
45		106.6	24.3	30000	6	AX250263	AX250263 Sequence

ALIGNMENTS

RESULT 1
AX006889
LOCUS AX006889 12381 bp DNA linear PAT 06-SEP-2000
DEFINITION Sequence 1 from Patent WO0001827.
ACCESSION AX006889
VERSION AX006889.1 GI:9994904
KEYWORDS .
SOURCE Streptomyces avermitilis.
ORGANISM Streptomyces avermitilis
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE 1 (bases 1 to 12381)
AUTHORS Kellenberger,J.L., Leadlay,P.F., Staunton,J., Mcarthur,H.A. and
Stutzman-Engwall,K.J.
TITLE Polyketides, their preparation, and materials for use therein

```
  JOURNAL   Patent: WO 0001827-A 1 13-JAN-2000;
            KELLENBERGER JOHANNES LAURENZ (CH); LEADLAY PETER FRANCIS (GB);
            STAUNTON JAMES (GB); BIOTICA TECH LTD (GB); PFIZER (US); MCARTHUR
            HAMISH ALASTAIR IRVIN (US); STUTZMAN ENGWALL KIM JONELLE (US)
FEATURES             Location/Qualifiers
     source          1. .12381
                     /organism="Streptomyces avermitilis"
                     /db_xref="taxon:33903"
BASE COUNT     1884 a   4561 c   4005 g   1931 t
ORIGIN

  Query Match             100.0%;  Score 438;  DB 6;  Length 12381;
  Best Local Similarity   100.0%;  Pred. No. 4.8e-72;
  Matches  438;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 TACTGGCTCGAAAGCACACAGCCCGGTGCCGGCAACGTGTCAGCAGCCGGACTCGACCCC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9764 TACTGGCTCGAAAGCACACAGCCCGGTGCCGGCAACGTGTCAGCAGCCGGACTCGACCCC 9823

Qy         61 ACCGAACACCCCCTACTCGGCGCCACATTGGAACTGGCGACTGACGGTGGAGCGCTTCTT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9824 ACCGAACACCCCCTACTCGGCGCCACATTGGAACTGGCGACTGACGGTGGAGCGCTTCTT 9883

Qy        121 GCAGGGCGCTTGTCTTTGAGGTCGCATCCGTGGCTGGCTGACCATGCCGTCGGCGGCACG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9884 GCAGGGCGCTTGTCTTTGAGGTCGCATCCGTGGCTGGCTGACCATGCCGTCGGCGGCACG 9943

Qy        181 GTGCTGCTGTCGGGCGCCACCTTCCTCGAACTCGCCCTTCATGCGGGCACATACGTGGGC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9944 GTGCTGCTGTCGGGCGCCACCTTCCTCGAACTCGCCCTTCATGCGGGCACATACGTGGGC 10003

Qy        241 TGCGACCGAGTGGATGAGCTGACGCTGCATGCGCCGCTGGTGGTTCCTGTGGATGGGGGT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10004 TGCGACCGAGTGGATGAGCTGACGCTGCATGCGCCGCTGGTGGTTCCTGTGGATGGGGGT 10063

Qy        301 GTGAGTGTGCAGGTTGGGGTTGCGGCTGCGGATGGGGAGGGGCGGCGTTTGGTGAGTGTG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10064 GTGAGTGTGCAGGTTGGGGTTGCGGCTGCGGATGGGGAGGGGCGGCGTTTGGTGAGTGTG 10123

Qy        361 TATGCGCGGGGGTGGGAGTGCTTGTGGTGGGGGTGGTGCGTCGGGTGGGGTGTGGACGTGT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10124 TATGCGCGGGGTGGGAGTGCTTGTGGTGGGGGTGGTGCGTCGGGTGGGGTGTGGACGTGT 10183

Qy        421 CATGCCTCGGGGGTGCTG 438
              ||||||||||||||||||
Db      10184 CATGCCTCGGGGGTGCTG 10201


RESULT 2
E38020
LOCUS       E38020                30690 bp    DNA     linear   PAT 31-JAN-2002
DEFINITION  Avermectin aglycon synthase gene.
ACCESSION   E38020
VERSION     E38020.1  GI:18626909
KEYWORDS    JP 2000245457-A/1.
SOURCE      Streptomyces avermitilis.
  ORGANISM  Streptomyces avermitilis
            Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
            Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1  (bases 1 to 30690)
  AUTHORS   Omura,S. and Ikeda,H.
  TITLE     Avermectin aglycon synthase gene
  JOURNAL   Patent: JP 2000245457-A 1 12-SEP-2000;
            THE KITASATO INSTITUTE
COMMENT     OS   Streptomyces avermitilis
            PN   JP 2000245457-A/1
            PD   12-SEP-2000
            PF   24-FEB-1999 JP 1999046961
            PR
            PI   SATOSHI OMURA,HARUO IKEDA
            PC   C12N15/00,A61K31/70,C12N1/15,C12N9/88,C12P19/62, PC
            C12Q1/68//C07D493/22,
            PC   (C12N1/15,C12R1:465),(C12N9/88,C12R1:465),C12N15/00 CC
            FH   Key             Location/Qualifiers
            FT   CDS             (1). .(11916)
            FT   CDS             (11971). .(30687).
FEATURES             Location/Qualifiers
     source          1. .30690
                     /organism="Streptomyces avermitilis"
                     /db_xref="taxon:33903"
BASE COUNT     5355 a  12455 c   8617 g   4263 t
ORIGIN

  Query Match             100.0%;  Score 438;  DB 6;  Length 30690;
  Best Local Similarity   100.0%;  Pred. No. 3.9e-72;
  Matches  438;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 TACTGGCTCGAAAGCACACAGCCCGGTGCCGGCAACGTGTCAGCAGCCGGACTCGACCCC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8947 TACTGGCTCGAAAGCACACAGCCCGGTGCCGGCAACGTGTCAGCAGCCGGACTCGACCCC 9006

Qy         61 ACCGAACACCCCCTACTCGGCGCCACATTGGAACTGGCGACTGACGGTGGAGCGCTTCTT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9007 ACCGAACACCCCCTACTCGGCGCCACATTGGAACTGGCGACTGACGGTGGAGCGCTTCTT 9066

Qy        121 GCAGGGCGCTTGTCTTTGAGGTCGCATCCGTGGCTGGCTGACCATGCCGTCGGCGGCACG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9067 GCAGGGCGCTTGTCTTTGAGGTCGCATCCGTGGCTGGCTGACCATGCCGTCGGCGGCACG 9126

Qy        181 GTGCTGCTGTCGGGCGCCACCTTCCTCGAACTCGCCCTTCATGCGGGCACATACGTGGGC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9127 GTGCTGCTGTCGGGCGCCACCTTCCTCGAACTCGCCCTTCATGCGGGCACATACGTGGGC 9186

Qy        241 TGCGACCGAGTGGATGAGCTGACGCTGCATGCGCCGCTGGTGGTTCCTGTGGATGGGGGT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9187 TGCGACCGAGTGGATGAGCTGACGCTGCATGCGCCGCTGGTGGTTCCTGTGGATGGGGGT 9246

Qy        301 GTGAGTGTGCAGGTTGGGGTTGCGGCTGCGGATGGGGAGGGGCGGCGTTTGGTGAGTGTG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9247 GTGAGTGTGCAGGTTGGGGTTGCGGCTGCGGATGGGGAGGGGCGGCGTTTGGTGAGTGTG 9306

Qy        361 TATGCGCGGGGGTGGGAGTGCTTGTGGTGGGGGTGGTGCGTCGGGTGGGGTGTGGACGTGT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9307 TATGCGCGGGGTGGGAGTGCTTGTGGTGGGGGTGGTGCGTCGGGTGGGGTGTGGACGTGT 9366

Qy        421 CATGCCTCGGGGGTGCTG 438
              ||||||||||||||||||
Db       9367 CATGCCTCGGGGGTGCTG 9384


RESULT 3
AB032367
LOCUS       AB032367               64957 bp    DNA     linear   BCT 18-SEP-1999
DEFINITION  Streptomyces avermitilis polyketide synthase gene cluster (aveA1,
            aveA2, aveA3, aveA4) and aveC, aveE genes, complete cds.
ACCESSION   AB032367
VERSION     AB032367.1  GI:5902890
KEYWORDS    AveC; cytochrome P450 hydroxylase; type I polyketide synthase AVES
            4; type I polyketide synthase AVES 3; type I polyketide synthase
            AVES 2; type I polyketide synthase AVES 1.
SOURCE      Streptomyces avermitilis DNA.
  ORGANISM  Streptomyces avermitilis
            Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
            Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1  (bases 1 to 64957)
  AUTHORS   Ikeda,H., Nonomiya,T., Usami,M., Ohta,T. and Omura,S.
  TITLE     Organization of the biosynthetic gene cluster for the polyketide
            anthelmintic macrolide avermectin in Streptomyces avermitilis
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9509-9514 (1999)
  MEDLINE   99380548
REFERENCE   2  (bases 1 to 64957)
  AUTHORS   Ikeda,H., Nonomiya,T., Usami,M., Ohta,T. and Omura,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-SEP-1999) Haruo Ikeda, School of Pharmaceutical
            Sciences, Kitasato University, Microbial chemistry; 5-9-1
            Shirokane, Minato-ku, Tokyo 108-8641, Japan
```

(E-mail:ikeda@mc.pharm.kitasato-u.ac.jp, Tel:+81-3-5791-6242,
Fax:+81-3-3444-6197)

FEATURES Location/Qualifiers

source 1. .64957
/organism="Streptomyces avermitilis"
/db_xref="taxon:33903"

gene 101. .12019
/gene="aveA1"

CDS 101. .12019
/gene="aveA1"
/note="contains a loading module, Ave module 1 and module 2
multifunctional polyketide synthase"
/codon_start=1
/transl_table=11
/product="type I polyketide synthase AVES 1"
/protein_id="BAA84474.1"
/db_xref="GI:5902891"
/translation="MQRMDGGEEPRPAAGEVLGVADEADGGVVFVFPGQGPQWPGMGR
ELLDASDVFRESVRACEAAFAPYVDWSVEQVLRDSPDAPGLDRVDVVQPTLFAVMISL
AALWRSQGVEPCAVLGHSLGEIAAAHVSGGLSLADAARVVTLWSQAQTTLAGTGALVS
VAATPDELLPRIAPWTEDNPARLAVAAVNGPRSTVVSGAREAVADLVADLTAAQVRTR
MIPVDVPAHSPLMYAIEERVVSGLLPITPRPSRIPFHSSVTGGRLDTRELDAAYWYRN
MSSTVRFEPAARLLLQQGPKTFVEMSPHPVLTMGLQELAPDLGDTTGTADTVIMGTLR
RGQGTLDHFLTSLAQLRGHGETSATTVLSARLTALSPTQQQSLLLDLVRAHTMAVLND
DGNERTASDAGPSASFAHLGFDSVMGVELRNRLSKATGLRLPVTLIFDHTTPAAVAAR
LRTAALGHLDEDTAPVPDSPSGHGGTAAADDPIAIIGMACRFPGGVRSPKDLWELAAS
GGDAIGPFPTDRGWPTEQRHAQDPTQPGTFYPQGGGFLHDAAHFDAGFFGISPREALA
MDPQQRLLLETSWEAFERAGIDPLSVRGSRTGVFAGALSFDYGPRMDTASSEGAADVE
GHILTGTTGSVLSGRIAYSFGLEGPAITVDTGCSASLVTLHLACQSLRSGECTLALAG
GVSVMSTLGMFIEFSRQRGLSVDGRCKAYSAAADGTGWGEGVGMLLVERLSDAVRLGH
RVLAVVRGSAVNQDGASNGLTAPNGPAQERVIRQALANAGLSVADVDVVEGHGTGTTL
GDPIEAQALLATYGQRAGDRPLWLGSLKSNIGHTMAAAGVGGVIKMVMALREGVLPRT
LHVDKPSPQVDWSAGAVRLLTEAVPWPGDAAGRLRRAGVSSFGIGGTNAHVILEEAPA
AGGCVAGGGVLEGAPGLAISVAESVAAPVAVSAPVAESVPVPVPVPVPVSARSEAG
LRAQAEALRQYVAVRPDVSLADVGAGLACGRAVLEHRAVVLAADREELVQGLGALAAG
EPDRRVTTGHAPGGDRGGVVFVFPGQGGQWAGMGVRLLASSPVFARRMQACEEALAPW
VDWSVVDILRRDAGDAVWERADVVQPVLFSVMVSLAALWRSYGIEPDAVLGHSQGEIA
AAHVCGALSLKDAAKTVALRSRALAAVRGRGGMASVPLPAQEVEQLIGERWAGRLWVA
AVNGPRSTAVSGDAEAVDEVLAYCAGTGVRARRIPVDYASHCPHVQPLREELLELLGD
ISPQPSGVPFFSTVEGTWLDTTTLDAAYWYRNLHQPVRFSDAVQALADDGHRVFVEVS
PHPTLVPAIEDTTEDTAEDVTAIGSLRRGDNDTRRFLTALAHTHTTGIGTPTTWHHHY
THHHTHPHPHTHLDLPTYPFQHQHYWLESSQPGAGSGSGAGAGSGAGSGRAGTAGGTA
EVESRFWDAVARQDLETVATTLAVPPSAGLDTVVPALSAWHRHQHDQARINTWTYQET
WKPLTLPTTHQPHQTWLIAIPETQTHHPHITNILTNLHHHGITPIPLTLNHTHTNPQH
LHHTLHHTRQQAQNHTTGAITGLLSLLALDETPHPHHPHTPTGTLLNLTLTQTHTQTH
PPTPLWYATTNATTTHPNDPLTHPTQAQTWGLARTTLLEHPTHTAGIIDLPTTPTPHT
LQHLTQTLTQPHHQTQLAIRTTGTHTRRLTPTTLTPTHQPPTPTPHGTTLITGGTGAL
ATHLTHHLTTHQPTQHLLLTSRTGPHTPHAQHLTTQLQQKGIHLTITTCDTSNPDQLQ
QLLNTIPPQHPLTTVIHTAGILDDATLTNLTPTQLNNVLRAKAHSAHLLHQLTQHTPL
TAFVLYSSAAATFGAPGQANYAAANAYLDALAHHRHTHHLPATSIAWGTWQGNGLADS
DKARAYLDRRGFRPMSPELATAAVTQAIADTERPYVVIADIDWSKIEHTSQTSDLVSA
ARELREPAVQRPTPPAELHKTLAHQTSADQRAALLELVRDHVAAVLRHADPKAIAPDQS
FRALGFDSLTAVEFRNLLIKATGLRLPVSLVFDHPTPAKLAVHLQNQLRGTAAESAPS
AAAVTAEASVTEPIAIVGMACRFPGGVTSADDFWDLISSEQDAIGGFPTDRGWDLDTL
YDPDPDHPGTCYTRNGGFLYDAGHFDAEFFGISPREALAMDPQQRLLLETAWETIEHA
GINPHTLHGTPTGVFTGTNGQDYALRVHNAGQSTDGFALTGTAGSVISGRISYTFGFE
GPAVSVDTACSSSLVALHLACQALRAGECSMALAGGVTVMSSPGAFVEFSRQRGLAAD
GHCKAFSAAADGTGWGEGVGMLLVERLSDAHRNGHRVLAVVRGSAVNQDGASNGLTAP
NGPSQQRVIRQALANAGLSAGDVDAVEAHGTGTTLGDPIEAQALLATYGQDRAGEGPL
WLGSVKSNVGHTQAAAGVAGVIKMVMALRHGLLPRTLHVDEPSPHVDWSAGAVQLLTE
TVPWPGGEGRLRRAGVSSFGVSGTNAHVILEEAPADDVPGGPPAGEGDAGSDDEAAAG
SPGVWPWLVSAKSQPALRAQAQALHAHLTDHPGLDLADVGYTLAHARAVFDHRATLIA
ADRDTFLQALQALAAGEPHPAVIHSSAPGGTGTGEAAGKTAFICSGQGTQRPGMAHGL
YHTHPVFAAALNDICTHLDPHLDHPLLPLLTQNDNDNEDAAALLQQTRYAQPALFAFQ
VALHRLLTDGYHITPHYYAGHSLGEITAAHLAGILTLTDATTLITQRATLMQTMPPGT
MTTLHTTPHHITHHLTAHENDLAIAAINTPTSLVISGTPHTVQHITTLCQQQGIKTKT
LPTNHAFHSPHTNPILNQLHQHTQTLTYHPPHTPLITANTPPDQLLTPHYWTQQARNT
VDYATTTQTLHQHGVTTYIELGPDNTLTTLTHHNLPNPPTTTLTLTHPHHHPQTHLLT
NLAKTTTTWHPHHYTHHDNQPHTHTHLDLPTYPFQHHHYWLESTQPGAGNVSAAGLDP
TEHPLLGATLELATDGGALLAGRLSLRSHPWLADHAVGGTVLLSGATFLELALHAGTY
VGCDRVDELTLHAPLVVPVDGGVSVQVGVAAADGEGRRLVSVYARGGSACGGGASGG
VWTCHASGVLVEAAAGGVVVDGLAGVWPPRGAVAVDVDGVRDRLAGAGCVLGPVFSGL
RAVWRDGGDLLAEVCLPEEAWGDAAGFGLHPALLDGVVQPLSVLLPGGTGFGEGAGFG
EGVRVPAVWGGVSLHRAGVTGVRVRVSAVGRGGGREAVSVVVGDEAGVPVASVDRLEL
RPVDMGQLRAVSVSAGRRGSLYAVQWAEVGPVPVCGQAWAWHEDVGESGGGPVPGVVV
LRCPDAGAGGGGGGGGGVGEVVGGVLGVVQGWLGLERPAGSRLVVVTRGAVVAGPE
DGPVDVVGASVWGLVRSAQAEHPDRFVLLDLDTDTGTDLDTGAGAGWGVDGGRVAAVV
ACGEPQLAVRGERLLAARLKRLESSGDVPAQRSGDTRARRSDVPAQRSGGVPARRSVD
VSGREVLPWLSGGSVLVTGGTGVLGAAVARHLAGVCGVRDLLLVSRRGPDAPGAEGLR
AELAALGAEVRIVACDVGERREVVRLLEGVPAGCPLTGVVHAAGVLDDATIASLTPER
LGTVFAAKVDAALLLDELTRGMELSAFVLFSSAAGILGSAGQGNYAAANAALDALAYR
RRAAGLPGVSLAWGLWEEASGMTGHLAGTDHRRIIRSGLHPMSTPDALALFDAALALD
RPVLLPADLRPAPPLPPLLQDLLPATRRRTTRTTTTGGADNGAQLHARLAGQTHEQQH
TTLLALVRSHIATVLGHTTPDTIPPDRAFRDLGFDSLTAVELRNRLSRTTGLRLPTTL
AFDHPNPTTLTHHLHTQLQPQPDNAVAPVLAELDKLESALSALDKTDSASERVTLRLK
SLMLRWNAPQHPTAESADDDEKFTSATEAEIFKFIDNDLGLS"

gene 12071. .30790
/gene="aveA2"

CDS 12071. .30790
/gene="aveA2"
/note="contains Ave module 3, module 4, module 5 and module 6
multifunctional polyketide synthase"
/codon_start=1
/transl_table=11
/product="type I polyketide synthase AVES 2"
/protein_id="BAA84475.1"
/db_xref="GI:5902892"
/translation="MQLANEAKLLEYLKRVTADLDRTRRRLYEVVEREQEPIAIVGMA
CRYPGGATSPTRLWHLVKSQTDAIGEFPTDRGWNLEQLYDPDPDRSGTSYTRSGGFLY
DAGDFDAAFFELSPREALAMDPQQRLLLETTWETFEQGGIDPRSMRGSRTGVFVGINP
EDYTTGYTHQPSNAVEGYLLTGSAASIASGRISYNFGLEGPAITIDTACSSSLVALHL
ACQALRSGECTMALAGGASVMATPFVFTEFSRQRGLAADGRCKAFSAAADGTGWSEGV
GMLLVERLSDARRNGHRVLAVVRGSAVNQDGASNGLTAPNGRSQVKVIRQALANAHLS
PADVDAVEAHGTGTTLGDPIEAQALVEAYGQDRPNGRPLWLGTLKSNIGHSMAAAGVG
GVIKMVMALRNGLLPRTLHVDEPSPHVDWSAGAVQLLTETVPWPGGEGRLRRAGVSSF
GVSGTNAHVILEEAPAHNIPSDTPADDVPGESAADEDAGSGDEAAAGSPGVWPWLVSA
KSQPALRAQAQALHAHLTDHPGLDLADVGYTLAHARAVFDHRATLIAADRDTFLQALQ
ALAAGEPHPAVIHSSAPGGTGTGEAAGKTAFICSGQGTQRPGMAHGLYHTHPVFAAAL
NDICTHLDPHLDHPLLPLLTQDPNTQDTTTLEEAAALLQQTRYAQPALFAFQVALHRL
LTDGYHITPHYYAGHSLGEITAAHLAGILTLTDATTLITQRATLMQTMPPGTMTTLHT
TPHHITHHLTAHENDLAIAAINTPTSLVISGTPHTVQHITTLCQQQGIKTKTLPTNHA
FHSPHTNPILNQLHQHTQTLTYHPPHTPLITANTPPDQLLTPHYWTQQARNTVDYATT
TQTLHQHGVTTYIELGPDNTLTTLTHDNLPNTPTTTLTLTHPHHHPQTHLLTNLAKTT
TTWHPHHYTHHHNQPHTHTHLDLPTYPFQHHHYWLQPPGKPSDPSPSEGREQATTPST
PLRDVLVGKSPQERDEELLRLVRTHAAAVLGHATPEVIVPNKAFKELGFDSLAAIQLR
NRLLADVDLPLPATLIFDYPTPMALCQFLRAAIVGADTGTTTRLPLTAVPADEPIAIV
GMACRYPGDVRTVDDLWQVVSGGHDAIGGFPTNRGWDLDTLYNPDPDHHGTSYTRSGG
FLYDAGNFDPDFFGISPREALAMDPQQRLLLETAWESIEHACINPDSLRGTPTGVFAG
LTYHDYAARFPTAPAGFEGYLGHGSAGSIASGRVAYALGLEGPALTVDTACSSSLVAL
HLACQALRSGECSMALAGGVTVMSTPAGPVEFSRQRGLAVDGRCKAFSAAADGTGWGE
GVGMLLVERLSDARRLGHRILAVVRGSAVNQDGASNGLTAPNGPSQERVIRLALANAG
LTPADVDAVEAHGTGTTLGDPIEAQALLATYGQDRPGNEPLWLGSMKSNIGHAQAAAG
VGGVIKMVMALRNGLLPRTLHVDEPSPHVDWSAGAVQLLTETVPWPGGEGRLRRAGVS
SFGVSGTNAHVILEEAPAHNIPSDTPADDAPGEAAADDVPGEAAGDDAGTGGEATGPA
AGSPGVWPWLVSAKSQPALRAQAQALHAHLTDHPGLDLADVGYTLAHARAVFDHRATL
IAADRDTFLQALQALAAGEPHPAVIHSSAPGGTGTGEAAGKTAFICSGQGTQRPGMAH
GLYHTHPVFAAALNDICTHLDPHLDHPLLPLLTQDPNTQDTTTLEEAAALLQQTPYAQ
PALFAFQVALHRLLTDGYHITPHYYAGHSLGEITAAHLAGILTLTDATTLITQRATLM
QTMPPGTMTTLHTTPHHITHHLTAHENDLAIAAINTPTSLVISGTPHTVQHITTLCQQ
QGIKTKTLPTKNAPHSPHTNPILNQLHQHTQTLTYHPPHTPLITANTPPDQLLTPHYW
TQQARNTVDYATTTQTLHQHGVTTYIELGPDNTLTTLTHHNLPNTPTTTLTLTHPHHH
PQTHLLTNLAKTTTTWHPHHYTHHHNQPHTHTHLDLPTYPFQHQHYWLESTQPGAGSG
SGSGSGRAGTAGGTAEVESRFWDAVARQDLETVATTLAVPPSAGLDTVVPALSAWHRH
QHDQARINTWTYQETWKPLTLPTTHQPHQTWLIAIPETQTHHPHITNILTNLHHHGIT
PIPLTLNHTHTNPQHLHHTRQQAQNHTTGPITGLLSLLALDETPHPHHPHTPTGTLLN
LTLTQTHTQTHPPTPLWYATTNATTTHPNDPLTHPTQAQTWGLARTTLLEHPTHTAGI
IDLPTTPTPHTLHHLTQTLTQPHHQTQLAIRTTGTHTRRLTPTTLTPTHQPPTPTPHG
TTLITGGTGALATHLTHHLTTHQPTQHLLLTSRTGPHTPHAQHLTTQLQQKGIHLTIT
TCDTSNPDQLQQLLNTIPPQHPLTTVIHTAGILDDATLTNLTPTQLNNVLRAKAHSAH
LLHQLTQHTPLNAFVLYSSAAATFGAPGQANYAAANAYLDALAHHRHTHHLPATSIAW
GTWQGNGLATGQVSEHLRRRGMFAMPPELAVTAVDGAIASGRPSLLVADIDWKKLGPV
LSSKSSVLLEDLPQAQGTEEARSTVEQTESTNLRQLLMGRSRSEQEEELLSLVRIHSA
AVLGRDDSEAIPPGRLFRDLGFDSLAAVELRNHLAAQTELALPTTLVFDYPSPTKLAQ
FLLSEIAEFQPDNSTPLPRPRAELDEPIAIVGMACRFPGGVTSADDFWDLISSEQDAI

GGFPTDRGWDLDTLYDPDPDHPGTCYTRNGGFLYDAGHFDAEFFGISPREALAMDPQQ
RLLLETAWETIEHAGINPHTLHGTPTGVFTGTNGQDHAAHIRQAPSGTEGFVLTGAAT
SIASGRISYILGLEGPAVTLDTACSSSLVALHLACQSLRSGECTMALAGGATVMTTPI
TFTEFARQRGLAPDGRCKAFSAAADGTGWGEGVGMLLVERLSDARRNGHRVLAVVRGS
AVNQDGASNGLTAPNGPSQQRVIRQALANADLTPADVDAVEAHGTGTTLGDPIEAQAI
LATYGQDRPGNGPLWLGSVKSNVGHTQAAAGVAGVIKMVMALRHRTLPPTLHADEPSP
HVDWSAGAVQLLTETVPWPGGEGRPRRAGVSSFGVSGTNAHVILEEAPADDVPGGPPA
DEDAGSGEEAAAGSPGVWPWLVSAKSQPALRAQAQALHAHLTDHPGLDLADVGYTLAH
ARAVFDHRATLIAADRDTFLQALQALAAGEPHPAVIHSSAPGGTGTGEAAGKTAFICS
GQGTQRPGMAHGLYHTHPVFAAALNDICTHLDPHLDHPLLPLLTQNDNDNDNEDAAAL
LQQTPYAQPALFAFQVALHRLLTDGYHITPHYYAGHSLGEITAAHLAGILTLTDATTL
ITQRATLMQTMPPGTMTTLHTTPHHITHHLTAHENDLAIAAINTPTSLVISGTPHTVQ
HITTLCQQQGIKTKTLPTNHAFHSPHTNPILNQLHQHTQTLTYHPPHTPLITANTPPD
QLLTPHYWTQQARNTVDYATTTQTLHQHGVTTYIELGPDNTLTTLTHHNLPNTPTTTL
TLTHPHHHPQTHLLTNLAKTTTTWHPHHYTHHHNQPHTHTHLDLPTYPFQHHHYWLEL
PSAQTSPGQRRSRRSAPDTAESEFWDAVNEEDLQSLAETLDIDASALDTVVPALSAWH
RHQHDQARINTWTYQETWKPLTLPTTHQPHQTWLIAIPETQTHHPHITNILTNLHHHG
ITPIPLTVNHTHTNPQHLHHTLHHTRQQAQNHTTGPITGLLSLLALDETPHPHHPHTP
TGTLLNLTLPQTHTQTHPPTPLWYATTNATTTHPNDPLTHPTQAQTWGLARTTLLEHP
THTAGIIDLPTTPTPHTLHHLTQTLTQPHHQTQLAIRTTGTHTRRLTPTTLTPTHQPP
TPTPHGTTLITGGTGALATHLTHHLTTHQPTQHLLLTSRTGPHTPHAQHLTTQLQQKG
IHLTITTCDTSNPDQLQQLLNTIPPQHPLTTVIHTAGVNLFAPVSETDAESFSSVTAA
KATGAAILHELLLDHETLEHFILFSSGAGAWGSGNQCAYSAANAYLDALATHRQTHGL
PGASIAWGPWAGKGMSAGDAAHGYLEKRGILPMEPRMALAAFHRARAQRPNSNLIIAD
IDWERFVPAFTARRHSPLIEDIPEVRQAAQELEAAASTAKTTTAQPIATSLRERLARL
TSSKQNQVLLGLIRTGICTVLGLRNPEGIEDQRAFRDLGFDSLTSAQFSKELAKETGL
PLPPSLVFDYPTPQECAAHLRTQLVDLDDEEDAALSNALPQVAHRRTVEDEPIAIIGM
ACRFPGGVRSADDLWELLASGKDAIGVFPTDRGWDLDTLYDPDPDHPGTCYTRNGGFL
YGAGHFDAEFFGISPREALAMDPQQRLLLETAWETIEHAGINPHTLHGTPTGVFAGIN
AQDHAAHIRQSRDVETIEGYALTGSSGSVASGRVAYTLGLEGPAVSVDTACSSSLVAL

Query Match 100.0%; Score 438; DB 1; Length 64957;
Best Local Similarity 100.0%; Pred. No. 3.4e-72;
Matches 438; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 TACTGGCTCGAAAGCACACAGCCCGGTGCCGGCAACGTGTCAGCAGCCGGACTCGACCCC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9047 TACTGGCTCGAAAGCACACAGCCCGGTGCCGGCAACGTGTCAGCAGCCGGACTCGACCCC 9106

Qy         61 ACCGAACACCCCCTACTCGGCGCCACATTGGAACTGGCGACTGACGGTGGAGCGCTTCTT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9107 ACCGAACACCCCCTACTCGGCGCCACATTGGAACTGGCGACTGACGGTGGAGCGCTTCTT 9166

Qy        121 GCAGGGCGCTTGTCTTTGAGGTCGCATCCGTGGCTGGCTGACCATGCCGTCGGCGGCACG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9167 GCAGGGCGCTTGTCTTTGAGGTCGCATCCGTGGCTGGCTGACCATGCCGTCGGCGGCACG 9226

Qy        181 GTGCTGCTGTCGGGCGCCACCTTCCTCGAACTCGCCCTTCATGCGGGCACATACGTGGGC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9227 GTGCTGCTGTCGGGCGCCACCTTCCTCGAACTCGCCCTTCATGCGGGCACATACGTGGGC 9286

Qy        241 TGCGACCGAGTGGATGAGCTGACGCTGCATGCGCCGCTGGTGGTTCCTGTGGATGGGGGT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9287 TGCGACCGAGTGGATGAGCTGACGCTGCATGCGCCGCTGGTGGTTCCTGTGGATGGGGGT 9346

Qy        301 GTGAGTGTGCAGGTTGGGGTTGCGGCTGCGGATGGGGAGGGGCGGCGTTTGGTGAGTGTG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9347 GTGAGTGTGCAGGTTGGGGTTGCGGCTGCGGATGGGGAGGGGCGGCGTTTGGTGAGTGTG 9406

Qy        361 TATGCGCGGGGTGGGAGTGCTTGTGGTGGGGGTGGTGCGTCGGGTGGGGTGTGGACGTGT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9407 TATGCGCGGGGTGGGAGTGCTTGTGGTGGGGGTGGTGCGTCGGGTGGGGTGTGGACGTGT 9466

Qy        421 CATGCCTCGGGGGTGCTG 438
              ||||||||||||||||||
Db       9467 CATGCCTCGGGGGTGCTG 9484
```

RESULT 4
AF275943
LOCUS AF275943 11096 bp DNA linear BCT 02-SEP-2000
DEFINITION Streptomyces avermitilis avermectin polyketide synthase gene, partial cds.

ACCESSION AF275943
VERSION AF275943.1 GI:9964075
KEYWORDS .
SOURCE Streptomyces avermitilis.
ORGANISM Streptomyces avermitilis
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae; Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE 1 (bases 1 to 11096)
AUTHORS Hong,Y.-S. and Lee,J.J.
TITLE Targeted Gene Disruption of the avermectin O-methyltransferase gene and polyketide synthase gene from Streptomyces avermitilis
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 11096)
AUTHORS Hong,Y.-S. and Lee,J.J.
TITLE Direct Submission
JOURNAL Submitted (07-JUN-2000) Anticancer Agent Research Laboratory, Korea Research Institute of Bioscience and Biotechnology, P.O. Box 116, YuSong-Gu, Taejon 305-600, South Korea
FEATURES Location/Qualifiers
source 1. .11096
/organism="Streptomyces avermitilis"
/strain="ATCC31271"
/db_xref="ATCC:31271"
/db_xref="taxon:33903"
CDS 220. .>11096
/note="Avr1"
/codon_start=1
/transl_table=11
/product="avermectin polyketide synthase"
/protein_id="AAG09812.1"
/db_xref="GI:9964076"
/translation="MADEADGGVVFVFPGQGPQWPGMGRELLDASDVFRESVRACEAE
FAPYVDWSVEQVLRDSPDAPGLDRVDVVQRTLFAVMISLAALWRSQGVEPCAVLGHSL
GEIAAAHVSGGLSLADAARVGDAWSQAQTTLAGTGALVSVAATPDELLPRIAPWTEDN
PARLAVAAVNGPRSTVVSGAREAVADLVADLTAAQVRTRMIPVDVPAHSPLMYAIEER
VVSGLLPITPRPSRIPFHSSVTGGRLDTRELDAAYWYRNMSSTVRFEPAARLLLQQGP
KTFVEMSPHPVLTMGLQELAADLGDTTGTADTVIMGTLRRGQGTLDHFLTSLAQLRGH
GETSATTVLSARLTALSPTQQQSLLLDLVRAHTMAVLNDDGNERTASDAGPSASFAHL
GFDSVMGVELRNRLSKATGLRLPVTLILDHTTPAAVAARLRTAALGHLDEDTAPVPDS
PSGHGGTAAADDPIAIIGMACRFPGGVRSPKDLWELAASGGDAIGPFPTDRGWPTEQR
HAQDPTQPGTFYPQGGGFLHDAAHFDAGFFGISPREALAMDPQQRLLLETSWEAFERA
GIDPLSVRGSRTGVFAGALSFDYGPRMDTASSEGAADVEGHILTGTTGSVLSGRSAYS
FGLEGPAITVDTGCSASLVTLHLACQSLRSGECTFALAGGVSDVHPGMFIEFSRQCGL
SVDGRCKAVSAAADGTGWGEGIGLLAVVRGSAVNQDGASNGLTAPNGPAQERVIRQAL
ANAGLSVADVDVVEGHGTGTTLGDPIEAQALLATYGQRAGDRPLWLGSLKSNIGHTMA
AAGVGGVIKMVMALREGVLPRTLHVDEPSPQGLAAGAVRLLTEAVPWPGDAAGRLRRA
GVSSFGIGGTNAHVILEEAPAAGGCVAGGRVLEGAPGLAISVAESVAAPVAVSAPVAE
SVPVPVPVPVPVPVSARSEAGLRAQAEALRQYEAVQPDVSLADVGAGLACRQAVLEHH
VVILAACTSSSRSAARTTARSSSTARPQARPAPTSARRIGALAAGSGSAALTTGHAPG
GDRGGVVFVFPWQGGQWAGMGVRLLCLLRVFARRMQACEEALAPWVDWSVVDILRRDA
GDAVWEQADVVQPVLFSVMVSLAALWRSYGIEPNEVLGHSKDEIAAAHIYGALSLKDA
AKTVALPPQEVEQLIGERGGRLWVAAVNGPRSTAVSGDAEAVVEVLAYCAGTGVRPRI
PVDYASHCPHVQPLREELLELLGDISPQPYGVPFFSTVEGTWLDTTTLDAAYWYRNLH
QPVRFSHDVQALADQGHRVLLEVSPHPTLVPAIEDTTEDTARRHCDRQPPPRRERHP
LLPQRLRLDPYYRHRQTHHVAPLLDPPRHLPPPLDAPRPAHLPFQHQHYWLESSQPGA
GSGSGAGAGSGAGSGRAGTAGGTAEVESRFWDAVARQDLETVATTLAVPPSAGLDTVV
PALSAWHRHQHDQARINTWTYQETWKPLTLPTTHQPHQTWLIAIPETQTHHPHITNIL
TNLHHHGITPIPLPLTTHHTNPQHLHHTLHHTRQQAQNHTTGAITGLLSLLALDETPH
PHHPHTPTGTLLNLTLTQTHTQTHPPTPLWYATTNATTTHPNDPLTHPTQAQTWGLAR
TTLLEHPTHTAGIIDLPTTPTPHTLHHLTQTLTQPHHQTQLAIRTTGTHTRRLTPTTL
TPTHQPPTPTPHGTTLITGGTGALATHLTHHLTTHQPTQHLLLTSRTGPHTPHAQHLT
TQLQQKGIHLTITTCDTSTPRPTHNNSLNTIPPQHPVTTVIHTGGILDDATLTNLTPT
QLNNVLRAKAHSAHLLHQLTQHTPLTAFVLYSSAAATFGAPGQANYAAANAYLDALAH
HRHTHHLPATSIAWGTWQGNGLADSDKARAYLDRRGFRPMSPELATAAVTQAIADTER
PYVVIADIDWSKIEHTSQTSDLVSAARAREPAVQRPTPPAELHKTLAHQTSADQRAAL
LELVRDHVAAVLRHADPKAIAPDQSFRALGFDSLTAVEFRNLLIKATGLRLPVSLVFD
HPTPAKLAVHLQNQLRGTAAESAPSAAAVTAEASVTEPIAIVGMACRFPGGVTSADDF
WDLISSEQDAIGGFPTDRGWDLDTLYDPDPDHPGTCYTRNGGFLYDAGHFDAEFFGIS
PREALAMDPQQRLLLETAWETIEHAGINPHTLHGTPTGVFTGTNGQDYALRVHNAGQS
TDGFALTGTAGSVISGRISYTFGFEGPAVSVDTACSSSLVALHLACQALRAGECSMAL
AGGVTVMSSPGAFVEFSRQRGLAADGHCKAFSAAADGTGWGEGVGMLLVERLSDAHRN
GHRVLAVVRGSAVNQDGASNGLTAPNGPSQQRVIRQALANAGLSAGDVDAVEAHGTGT
TLGDPIEAQALLATYGQDRAGEGPLWLGSVKSNVGHTQAAAGVAGVIKMVMALRHGLL

PRTLHVDEPSPHVDWSAGAVQLLTETVPWPGGEGRLRRAGVSSFGVSGTNAHVILEEA
PADDVPGGPPAGEGDAGSDDEAAAGSPGVWPWLVSAKSQPALRAQAQALHAHLTDHPG
LDLADVGYTLAHARAVFDHRATLIAADRDTFLQALQALAAGEPHPAVIHSSAPGGTGT
GEAAGKTAFICSGQGTQRPGMAHGLYHTHPVFAAALNDICTHLDPHLDHPLLPLLTQN
DNDNEDAAALLQQTRYAQPALFAFQVALHRLLTDGYHITPHYYAGHSLGEITGAHLAG
ILTLTDATTLITQRATLMQTMPPGTMTTLHTTPHHITHHLTAHENDLAIAAINTPTSL
VISGTPHTVQHITTLCQQQGIKTKTLPTNHAFHSPHTNPILNQLHQHTQTLTYHPPHT
PLITANTPPDQLLTPHYWTQQARNTVDYATTTQTLHQHGVTTYIELGPDNTLTTLTHH
NLPNTPTTTLTLTHPHHHPQTHLLTNLAKTTTWHPHHYTHHAQPTPHPHPLDLPTYP
FQHHHYWLESTQPGAGNVSAAGLDPTEHPLLGATLELATDGGALLAGRLSLRSHPWLA
DHAVGGTVLLSGATFLELALHAGTYVGCDRVDELTLHAPLVVPVDGGVSVQVGVAAAD
GKGRRLVSVYARGGSACGGGGASGGVWTCHASGVLVEAAAGGVVVDGLAGVWPPRGAV
AVDVDGVRDRLAGAGCVLGPVFSGLRAVWRDGGDLLAEVCLPEEAWGDAAGFGLHPAL
LNGVVQPLSVLLPGGTGFGKGAGFGKGVRVPAVWGGVSLHRAGVTGVRVRVSAVGRGG
GREAVSVVVGDEAGVPVASVDRLELRPVDMGQLRAVSVSAGRRGSLYAVQWAEVGPVP
VCGQAWAWHEDVGESGGGPVPGVVVLRCPDAGAGGGGGGGGGGVGEVVGGVLGVVQG
WLGLERFAGSRLVVVTRGAVVAGPEDGPVDVVGASVWGLVRSAQAEHPDRFVLLDLDT
DTGTDLDTGAGAGWGVDGGRVAAVVACGEPQLAVRGERLLAARLTRLESSGDVPAQRS
GDTRARRSDVPAQRSGGVPARRSVDVSGREVLPWLSGGSVLVTGGTGVLGAAVARHLA
GVCGVRDLLLVSRRGPDAPGAEGLRRSWPRGAEVRIVACDVGERREVVRLAGGCSCRV
SVFVDFSAVASVTVRLPVASDVRKEAAMAYATVEEFTDYLDPDP"

BASE COUNT 1756 a 4059 c 3528 g 1753 t
ORIGIN

Query Match 99.3%; Score 434.8; DB 1; Length 11096;
Best Local Similarity 99.5%; Pred. No. 1.9e-71;
Matches 436; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
Qy          1 TACTGGCTCGAAAGCACACAGCCCGGTGCCGGCAACGTGTCAGCAGCCGGACTCGACCCC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9067 TACTGGCTCGAAAGCACACAGCCCGGTGCCGGCAACGTGTCAGCAGCCGGACTCGACCCC 9126

Qy         61 ACCGAACACCCCCTACTCGGCGCCACATTGGAACTGGCGACTGACGGTGGAGCGCTTCTT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9127 ACCGAACACCCCCTACTCGGCGCCACATTGGAACTGGCGACTGACGGTGGAGCGCTTCTT 9186

Qy        121 GCAGGGCGCTTGTCTTTGAGGTCGCATCCGTGGCTGGCTGACCATGCCGTCGGCGGCACG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9187 GCAGGGCGCTTGTCTTTGAGGTCGCATCCGTGGCTGGCTGACCATGCCGTCGGCGGCACG 9246

Qy        181 GTGCTGCTGTCGGGCGCCACCTTCCTCGAACTCGCCCTTCATGCGGGCACATACGTGGGC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9247 GTGCTGCTGTCGGGCGCCACCTTCCTCGAACTCGCCCTTCATGCGGGCACATACGTGGGC 9306

Qy        241 TGCGACCGAGTGGATGAGCTGACGCTGCATGCGCCGCTGGTGGTTCCTGTGGATGGGGGT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9307 TGCGACCGAGTGGATGAGCTGACGCTGCATGCGCCGCTGGTGGTTCCTGTGGATGGGGGT 9366

Qy        301 GTGAGTGTGCAGGTTGGGGTTGCGGCTGCGGATGGGGAGGGGCGGCGTTTGGTGAGTGTG 360
              |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db       9367 GTGAGTGTGCAGGTTGGGGTTGCGGCTGCGGATGGGAAGGGGCGGCGTTTGGTGAGTGTG 9426

Qy        361 TATGCGCGGGGTGGGAGTGCTTGTGGTGGGGGTGGTGCGTCGGGTGGGGTGTGGACGTGT 420
              |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db       9427 TATGCGCGGGGTGGAAGTGCTTGTGGTGGGGGTGGTGCGTCGGGTGGGGTGTGGACGTGT 9486

Qy        421 CATGCCTCGGGGGTGCTG 438
              ||||||||||||||||||
Db       9487 CATGCCTCGGGGGTGCTG 9504
```

RESULT 5
E38022
LOCUS E38022 438 bp DNA linear PAT 31-JAN-2002
DEFINITION Avermectin aglycon synthase gene.
ACCESSION E38022
VERSION E38022.1 GI:18626911
KEYWORDS JP 2000245457-A/3.
SOURCE Streptomyces avermitilis.
ORGANISM Streptomyces avermitilis
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE 1 (bases 1 to 438)
AUTHORS Omura,S. and Ikeda,H.
TITLE Avermectin aglycon synthase gene
JOURNAL Patent: JP 2000245457-A 3 12-SEP-2000;
THE KITASATO INSTITUTE
COMMENT OS Streptomyces avermitilis
PN JP 2000245457-A/3
PD 12-SEP-2000
PF 24-FEB-1999 JP 1999046961
PR
PI SATOSHI OMURA,HARUO IKEDA
PC C12N15/00,A61K31/70,C12N1/15,C12N9/88,C12P19/62, PC
C12Q1/68//C07D493/22,
PC (C12N1/15,C12R1:465),(C12N9/88,C12R1:465),C12N15/00 CC
FH Key Location/Qualifiers
FT CDS (1). .(486).
FEATURES Location/Qualifiers
source 1. .438
/organism="Streptomyces avermitilis"
/db_xref="taxon:33903"
BASE COUNT 55 a 112 c 178 g 93 t
ORIGIN

Query Match 98.5%; Score 431.6; DB 6; Length 438;
Best Local Similarity 99.1%; Pred. No. 1.5e-70;
Matches 434; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

```
Qy          1 TACTGGCTCGAAAGCACACAGCCCGGTGCCGGCAACGTGTCAGCAGCCGGACTCGACCCC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 TACTGGCTCGAAAGCACACAGCCCGGTGCCGGCAACGTGTCAGCAGCCGGACTCGACCCC 60

Qy         61 ACCGAACACCCCCTACTCGGCGCCACATTGGAACTGGCGACTGACGGTGGAGCGCTTCTT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ACCGAACACCCCCTACTCGGCGCCACATTGGAACTGGCGACTGACGGTGGAGCGCTTCTT 120

Qy        121 GCAGGGCGCTTGTCTTTGAGGTCGCATCCGTGGCTGGCTGACCATGCCGTCGGCGGCACG 180
              |||||||||||||||||||||||||||||||||||||||||||| | |  ||||||||||||
Db        121 GCAGGGCGCTTGTCTTTGAGGTCGCATCCGTGGCTGGCTGACTACGAGGTCGGCGGCACG 180

Qy        181 GTGCTGCTGTCGGGCGCCACCTTCCTCGAACTCGCCCTTCATGCGGGCACATACGTGGGC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GTGCTGCTGTCGGGCGCCACCTTCCTCGAACTCGCCCTTCATGCGGGCACATACGTGGGC 240

Qy        241 TGCGACCGAGTGGATGAGCTGACGCTGCATGCGCCGCTGGTGGTTCCTGTGGATGGGGGT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TGCGACCGAGTGGATGAGCTGACGCTGCATGCGCCGCTGGTGGTTCCTGTGGATGGGGGT 300

Qy        301 GTGAGTGTGCAGGTTGGGGTTGCGGCTGCGGATGGGGAGGGGCGGCGTTTGGTGAGTGTG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GTGAGTGTGCAGGTTGGGGTTGCGGCTGCGGATGGGGAGGGGCGGCGTTTGGTGAGTGTG 360

Qy        361 TATGCGCGGGGTGGGAGTGCTTGTGGTGGGGGTGGTGCGTCGGGTGGGGTGTGGACGTGT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 TATGCGCGGGGTGGGAGTGCTTGTGGTGGGGGTGGTGCGTCGGGTGGGGTGTGGACGTGT 420

Qy        421 CATGCCTCGGGGGTGCTG 438
              ||||||||||||||||||
Db        421 CATGCCTCGGGGGTGCTG 438
```

RESULT 6
E38021
LOCUS E38021 31422 bp DNA linear PAT 31-JAN-2002
DEFINITION Avermectin aglycon synthase gene.
ACCESSION E38021
VERSION E38021.1 GI:18626910
KEYWORDS JP 2000245457-A/2.
SOURCE Streptomyces avermitilis.
ORGANISM Streptomyces avermitilis
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE 1 (bases 1 to 31422)
AUTHORS Omura,S. and Ikeda,H.
TITLE Avermectin aglycon synthase gene

```
  JOURNAL   Patent: JP 2000245457-A 2 12-SEP-2000;
            THE KITASATO INSTITUTE
COMMENT     OS   Streptomyces avermitilis
            PN   JP 2000245457-A/2
            PD   12-SEP-2000
            PF   24-FEB-1999 JP 1999046961
            PR
            PI   SATOSHI OMURA,HARUO IKEDA
            PC   C12N15/00,A61K31/70,C12N1/15,C12N9/88,C12P19/62, PC
            C12Q1/68//C07D493/22,
            PC   (C12N1/15,C12R1:465),(C12N9/88,C12R1:465),C12N15/00 CC
            FH   Key             Location/Qualifiers
            FT   CDS             (1). .(14643)
            FT   CDS             (14824). .(31419).
FEATURES             Location/Qualifiers
     source          1. .31422
                     /organism="Streptomyces avermitilis"
                     /db_xref="taxon:33903"
BASE COUNT     4136 a  10237 c  11677 g   5372 t
ORIGIN

  Query Match             62.5%;  Score 273.6;  DB 6;  Length 31422;
  Best Local Similarity   78.6%;  Pred. No. 1.2e-41;
  Matches  327;  Conservative    0;  Mismatches   89;  Indels    0;  Gaps    0;

Qy         23 CCGGTGCCGGCAACGTGTCAGCAGCCGGACTCGACCCCACCGAACACCCCCTACTCGGCG 82
              | || || ||  |||| | || || ||  | || ||  |||||||||| || ||||  |
Db      17705 CGGGAGCAGGTGACGTCGCCGCTGCTGGCTTGGAGCCGGCCGAACACCCTCTGCTCGCGG 17764

Qy         83 CCACATTGGAACTGGCGACTGACGGTGGAGCGCTTCTTGCAGGGCGCTTGTCTTTGAGGT 142
              | ||| |  |||| ||       |  ||    || ||  | || ||| |||| ||| | |
Db      17765 CAACAGTCCAACTCGCAGACACGGACGGCTGCCTACTGACGGGTCGCCTGTCCTTGCGCT 17824

Qy        143 CGCATCCGTGGCTGGCTGACCATGCCGTCGGCGGCACGGTGCTGCTGTCGGGCGCCACCT 202
              |||||||||||||||  ||  | |  || || ||  |||| ||||||||||  |  | |
Db      17825 CGCATCCGTGGCTGGGCGATTACGAGGTGGGGGGTGCGGTCCTGCTGTCGGGGTCGGCGT 17884

Qy        203 TCCTCGAACTCGCCCTTCATGCGGGCACATACGTGGGCTGCGACCGAGTGGATGAGCTGA 262
              || | || || ||  | || |  |||  |  |||||||||  |||| | ||  | || |
Db      17885 TCGTGGAGCTGGCGGTCCAGGTTGGCGAACGCGTGGGCTGCACCCGAATCGAGCAACTCA 17944

Qy        263 CGCTGCATGCGCCGCTGGTGGTTCCTGTGGATGGGGGTGTGAGTGTGCAGGTTGGGGTTG 322
              |  |||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db      17945 CTGTGCATGCGCCGCTGGTGGTTCCTGTGGGTGGGGGTGTGAGTGTGCAGGTTGGGGTTG 18004

Qy        323 CGGCTGCGGATGGGGAGGGGCGGCGTTTGGTGAGTGTGTATGCGCGGGGTGGGAGTGCTT 382
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      18005 CGGCTGCGGATGGGGAGGGGCGGCGTTTGGTGAGTGTGTATGCGCGGGGTGGGAGTGCTT 18064

Qy        383 GTGGTGGGGGTGGTGCGTCGGGTGGGGTGTGGACGTGTCATGCCTCGGGGGTGCTG 438
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      18065 GTGGTGGGGGTGGTGCGTCGGGTGGGGTGTGGACGTGTCATGCCTCGGGGGTGCTG 18120


RESULT 7
AB032367/c
LOCUS       AB032367               64957 bp    DNA     linear   BCT 18-SEP-1999
DEFINITION  Streptomyces avermitilis polyketide synthase gene cluster (aveA1,
            aveA2, aveA3, aveA4) and aveC, aveE genes, complete cds.
ACCESSION   AB032367
VERSION     AB032367.1  GI:5902890
KEYWORDS    AveC; cytochrome P450 hydroxylase; type I polyketide synthase AVES
            4; type I polyketide synthase AVES 3; type I polyketide synthase
            AVES 2; type I polyketide synthase AVES 1.
SOURCE      Streptomyces avermitilis DNA.
  ORGANISM  Streptomyces avermitilis
            Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
            Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1  (bases 1 to 64957)
  AUTHORS   Ikeda,H., Nonomiya,T., Usami,M., Ohta,T. and Omura,S.
  TITLE     Organization of the biosynthetic gene cluster for the polyketide
            anthelmintic macrolide avermectin in Streptomyces avermitilis
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9509-9514 (1999)
  MEDLINE   99380548
REFERENCE   2  (bases 1 to 64957)
  AUTHORS   Ikeda,H., Nonomiya,T., Usami,M., Ohta,T. and Omura,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-SEP-1999) Haruo Ikeda, School of Pharmaceutical
            Sciences, Kitasato University, Microbial chemistry; 5-9-1
            Shirokane, Minato-ku, Tokyo 108-8641, Japan
            (E-mail:ikeda@mc.pharm.kitasato-u.ac.jp, Tel:+81-3-5791-6242,
            Fax:+81-3-3444-6197)
FEATURES             Location/Qualifiers
     source          1. .64957
                     /organism="Streptomyces avermitilis"
                     /db_xref="taxon:33903"
     gene            101. .12019
                     /gene="aveA1"
     CDS             101. .12019
                     /gene="aveA1"
                     /note="contains a loading module, Ave module 1 and module
                     2
                     multifunctional polyketide synthase"
                     /codon_start=1
                     /transl_table=11
                     /product="type I polyketide synthase AVES 1"
                     /protein_id="BAA84474.1"
                     /db_xref="GI:5902891"
                     /translation="MQRMDGGEEPRPAAGEVLGVADEADGGVVFVFPGQGPQWPGMGR
                     ELLDASDVFRESVRACEAAFAPYVDWSVEQVLRDSPDAPGLDRVDVVQPTLFAVMISL
                     AALWRSQGVEPCAVLGHSLGEIAAAHVSGGLSLADAARVVTLWSQAQTTLAGTGALVS
                     VAATPDELLPRIAPWTEDNPARLAVAAVNGPRSTVVSGAREAVADLVADLTAAQVRTR
                     MIPVDVPAHSPLMYAIEERVVSGLLPITPRPSRIPFHSSVTGGRLDTRELDAAYWYRN
                     MSSTVRFEPAARLLLQQGPKTFVEMSPHPVLTMGLQELAPDLGDTTGTADTVIMGTLR
                     RGQGTLDHFLTSLAQLRGHGETSATTVLSARLTALSPTQQQSLLLDLVRAHTMAVLND
                     DGNERTASDAGPSASFAHLGFDSVMGVELRNRLSKATGLRLPVTLIFDHTTPAAVAAR
                     LRTAALGHLDEDTAPVPDSPSGHGGTAAADDPIAIIGMACRFPGGVRSPKDLWELAAS
                     GGDAIGPFPTDRGWPTEQRHAQDPTQPGTFYPQGGGFLHDAAHFDAGFFGISPREALA
                     MDPQQRLLLETSWEAFERAGIDPLSVRGSRTGVFAGALSFDYGPRMDTASSEGAADVE
                     GHILTGTTGSVLSGRIAYSFGLEGPAITVDTGCSASLVTLHLACQSLRSGECTLALAG
                     GVSVMSTLGMFIEFSRQRGLSVDGRCKAYSAAADGTGWGEGVGMLLVERLSDAVRLGH
                     RVLAVVRGSAVNQDGASNGLTAPNGPAQERVIRQALANAGLSVADVDVVEGHGTGTTL
                     GDPIEAQALLATYGQRAGDRPLWLGSLKSNIGHTMAAAGVGGVIKMVMALREGVLPRT
                     LHVDKPSPQVDWSAGAVRLLTEAVPWPGDAAGRLRRAGVSSFGIGGTNAHVILEEAPA
                     AGGCVAGGGVLEGAPGLAISVAESVAAPVAVSAPVAESVPVPVPVPVPVPVSARSEAG
                     LRAQAEALRQYVAVRPDVSLADVGAGLACGRAVLEHRAVVLAADREELVQGLGALAAG
                     EPDRRVTTGHAPGGDRGGVVFVFPGQGGQWAGMGVRLLASSPVFARRMQACEEALAPW
                     VDWSVVDILRRDAGDAVWERADVVQPVLFSVMVSLAALWRSYGIEPDAVLGHSQGEIA
                     AAHVCGALSLKDAAKTVALRSRALAAVRGRGGMASVPLPAQEVEQLIGERWAGRLWVA
                     AVNGPRSTAVSGDAEAVDEVLAYCAGTGVRARRIPVDYASHCPHVQPLREELLELLGD
                     ISPQPSGVPFFSTVEGTWLDTTTLDAAYWYRNLHQPVRFSDAVQALADDGHRVFVEVS
                     PHPTLVPAIEDTTEDTAEDVTAIGSLRRGDNDTRRFLTALAHTHTTGIGTPTTWHHHY
                     THHHTHPHHPHTHLDLPTYPFQHQHYWLESSQPGAGSGSGAGAGSGAGSGRAGTAGGTA
                     EVESRFWDAVARQDLETVATTLAVPPSAGLDTVVPALSAWHRHQHDQARINTWTYQET
                     WKPLTLPTTHQPHQTWLIAIPETQTHHPHITNILTNLHHHGITPIPLTLNHTHTNPQH
                     LHHTLHHTRQQAQNHTTGAITGLLSLLALDETPHPHHPHTPTGTLLNLTLTQTHTQTH
                     PPTPLWYATTNATTTHPNDPLTHPTQAQTWGLARTTLLEHPTHTAGIIDLPTTPTPHT
                     LQHLTQTLTQPHHQTQLAIRTTGTHTRRLTPTTLTPTHQPPTPTPHGTTLITGGTGAL
                     ATHLTHHLTTHQPTQHLLLTSRTGPHTPHAQHLTTQLQQKGIHLTITTCDTSNPDQLQ
                     QLLNTIPPQHPLTTVIHTAGILDDATLTNLTPTQLNNVLRAKAHSAHLLHQLTQHTPL
                     TAFVLYSSAAATFGAPGQANYAAANAYLDALAHHRHTHHLPATSIAWGTWQGNGLADS
                     DKARAYLDRRGFRPMSPELATAAVTQAIADTERPYVVIADIDWSKIEHTSQTSDLVSA
                     ARELREPAVQRPTPPAELHKTLAHQTSADQRAALLELVRDHVAAVLRHADPKAIAPDQS
                     FRALGFDSLTAVEFRNLLIKATGLRLPVSLVFDHPTPAKLAVHLQNQLRGTAAESAPS
                     AAAVTAEASVTEPIAIVGMACRFPGGVTSADDFWDLISSEQDAIGGFPTDRGWDLDTL
                     YDPDPDHPGTCYTRNGGFLYDAGHFDAEFFGISPREALAMDPQQRLLLETAWETIEHA
                     GINPHTLHGTPTGVFTGTNGQDYALRVHNAGQSTDGFALTGTAGSVISGRISYTFGFE
                     GPAVSVDTACSSSLVALHLACQALRAGECSMALAGGVTVMSSPGAFVEFSRQRGLAAD
                     GHCKAFSAAADGTGWGEGVGMLLVERLSDAHRNGHRVLAVVRGSAVNQDGASNGLTAP
                     NGPSQQRVIRQALANAGLSAGDVDAVEAHGTGTTLGDPIEAQALLATYGQDRAGEGPL
                     WLGSVKSNVGHTQAAAGVAGVIKMVMALRHGLLPRTLHVDEPSPHVDWSAGAVQLLTE
                     TVPWPGGEGRLRRAGVSSFGVSGTNAHVILEEAPADDVPGGPPAGEGDAGSDDEAAAG
                     SPGVWPWLVSAKSQPALRAQAQALHAHLTDHPGLDLADVGYTLAHARAVFDHRATLIA
                     ADRDTFLQALQALAAGEPHPAVIHSSAPGGTGTGEAAGKTAFICSGQGTQRPGMAHGL
                     YHTHPVFAAALNDICTHLDPHLDHPLLPLLTQNDNDNEDAAALLQQTRYAQPALFAFQ
                     VALHRLLTDGYHITPHYYAGHSLGEITAAHLAGILTLTDATTLITQRATLMQTMPPGT
```

MTTLHTTPHHITHHLTAHENDLAIAAINTPTSLVISGTPHTVQHITTLCQQQGIKTKT
LPTNHAFHSPHTNPILNQLHQHTQTLTYHPPHTPLITANTPPDQLLTPHYWTQQARNT
VDYATTTQTLHQHGVTTYIELGPDNTLTTLTHHNLPNPPTTTLTLTHPHHHPQTHLLT
NLAKTTTTWHPHHYTHHDNQPHTHTHLDLPTYPFQHHHYWLESTQPGAGNVSAAGLDP
TEHPLLGATLELATDGGALLAGRLSLRSHPWLADHAVGGTVLLSGATFLELALHAGTY
VGCDRVDELTLHAPLVVPVDGGVSVQVGVAAADGEGRRLVSVYARGGSACGGGASGG
VWTCHASGVLVEAAAGGVVVDGLAGVWPPRGAVAVDVDGVRDRLAGAGCVLGPVFSGL
RAVWRDGGDLLAEVCLPEEAWGDAAGFGLHPALLDGVVQPLSVLLPGGTGFGEGAGFG
EGVRVPAVWGGVSLHRAGVTGVRVRVSAVGRGGGREAVSVVVGDEAGVPVASVDRLEL
RPVDMGQLRAVSVSAGRRGSLYAVQWAEVGPVPVCGQAWAWHEDVGESGGGPVPGVVV
LRCPDAGAGGGGGGGGGGVGEVVGGVLGVVQGWLGLERFAGSRLVVVTRGAVVAGPE
DGPVDVVGASVWGLVRSAQAEHPDRFVLLDLDTDTGTDLDTGAGAGWGVDGGRVAAVV
ACGEPQLAVRGERLLAARLKRLESSGDVPAQRSGDTRARRSDVPAQRSGGVPARRSVD
VSGREVLPWLSGGGSVLVTGGTGVLGAAVARHLAGVCGVRDLLLVSRRGPDAPGAEGLR
AELAALGAEVRIVACDVGERREVVRLLEGVPAGCPLTGVVHAAGVLDDATIASLTPER
LGTVFAAKVDAALLLDELTRGMELSAFVLFSSAAGILGSAGQGNYAAANAALDALAYR
RRAAGLPGVSLAWGLWEEASGMTGHLAGTDHRRIIRSGLHPMSTPDALALFDAALALD
RPVLLPADLRPAPPLPPLLQDLLPATRRRTTRTTTTGGADNGAQLHARLAGQTHEQQH
TTLLALVRSHIATVLGHTTPDTIPPDRAFRDLGFDSLTAVELRNRLSRTTGLRLPTTL
AFDHPNPTTLTHHLHTQLQPQPDNAVAPVLAELDKLESALSALDKTDSASERVTLRLK
SLMLRWNAPQHPTAESADDDEKFTSATEAEIFKFIDNDLGLS"
gene 12071. .30790
/gene="aveA2"
CDS 12071. .30790
/gene="aveA2"
/note="contains Ave module 3, module 4, module 5 and
module 6
multifunctional polyketide synthase"
/codon_start=1
/transl_table=11
/product="type I polyketide synthase AVES 2"
/protein_id="BAA84475.1"
/db_xref="GI:5902892"
/translation="MQLANEAKLLEYLKRVTADLDRTRRRLYEVVEREQEPIAIVGMA
CRYPGGATSPTRLWHLVKSQTDAIGEFPTDRGWNLEQLYDPDPDRSGTSYTRSGGFLY
DAGDFDAAFFELSPREALAMDPQQRLLLETTWETFEQGGIDPRSMRGSRTGVFVGINP
EDYTTGYTHQPSNAVEGYLLTGSAASIASGRISYNFGLEGPAITIDTACSSSLVALHL
ACQALRSGECTMALAGGASVMATPFVFTEFSRQRGLAADGRCKAFSAAADGTGWSEGV
GMLLVERLSDARRNGHRVLAVVRGSAVNQDGASNGLTAPNGRSQVKVIRQALANAHLS
PADVDAVEAHGTGTTLGDPIEAQALVEAYGQDRPNGRPLWLGTLKSNIGHSMAAAGVG
GVIKMVMALRNGLLPRTLHVDEPSPHVDWSAGAVQLLTETVPWPGGEGRLRRAGVSSF
GVSGTNAHVILEEAPAHNIPSDTPADDVPGESAADEDAGSGDEAAAGSPGVWPWLVSA
KSQPALRAQAQALHAHLTDHPGLDLADVGYTLAHARAVFDHRATLIAADRDTFLQALQ
ALAAGEPHPAVIHSSAPGGTGTGEAAGKTAFICSGQGTQRPGMAHGLYHTHPVFAAAL
NDICTHLDPHLDHPLLPLLTQDPNTQDTTTLEEAAALLQQTRYAQPALFAFQVALHRL
LTDGYHITPHYYAGHSLGEITAAHLAGILTLTDATTLITQRATLMQTMPPGTMTTLHT
TPHHITHHLTAHENDLAIAAINTPTSLVISGTPHTVQHITTLCQQQGIKTKTLPTNHA
FHSPHTNPILNQLHQHTQTLTYHPPHTPLITANTPPDQLLTPHYWTQQARNTVDYATT
TQTLHQHGVTTYIELGPDNTLTTLTHDNLPNTPTTTLTLTHPHHHPQTHLLTNLAKTT
TTWHPHHYTHHHNQPHTHTHLDLPTYPFQHHHYWLQPPGKPSDPSPSEGREQATTPST
PLRDVLVGKSPQERDEELLRLVRTHAAAVLGHATPEVIVPNKAFKELGFDSLAAIQLR
NRLLADVDLPLPATLIFDYPTPMALCQFLRAAIVGADTGTTTRLPLTAVPADEPIAIV
GMACRYPGDVRTVDDLWQVVSGGHDAIGGFPTNRGWDLDTLYNPDPDHHGTSYTRSGG
FLYDAGNFDPDFFGISPREALAMDPQQRLLLETAWESIEHACINPDSLRGTPTGVFAG
LTYHDYAARFPTAPAGFEGYLGHGSAGSIASGRVAYALGLEGPALTVDTACSSSLVAL
HLACQALRSGECSMALAGGVTVMSTPAGFVEFSRQRGLAVDGRCKAFSAAADGTGWGE
GVGMLLVERLSDARRLGHRILAVVRGSAVNQDGASNGLTAPNGPSQERVIRLALANAD
LTPADVDAVEAHGTGTTLGDPIEAQALLATYGQDRPGNEPLWLGSMKSNIGHAQAAAG
VGGVIKMVMALRNGLLPRTLHVDEPSPHVDWSAGAVQLLTETVPWPGGEGRLRRAGVS
SFGVSGTNAHVILEEAPAHNIPSDTPADDAPGEAAADDVPGEAAGDDAGTGGEATGPA
AGSPGVWPWLVSAKSQPALRAQAQALHAHLTDHPGLDLADVGYTLAHARAVFDHRATL
IAADRDTFLQALQALAAGEPHPAVIHSSAPGGTGTGEAAGKTAFICSGQGTQRPGMAH
GLYHTHPVFAAALNDICTHLDPHLDHPLLPLLTQDPNTQDTTTLEEAAALLQQTPYAQ
PALFAFQVALHRLLTDGYHITPHYYAGHSLGEITAAHLAGILTLTDATTLITQRATLM
QTMPPGTMTTLHTTPHHITHHLTAHENDLAIAAINTPTSLVISGTPHTVQHITTLCQQ
QGIKTKTLPTKNAFHSPHTNPILNQLHQHTQTLTYHPPHTPLITANTPPDQLLTPHYW
TQQARNTVDYATTTQTLHQHGVTTYIELGPDNTLTTLTHHNLPNTPTTTLTLTHPHHH
PQTHLLTNLAKTTTTWHPHHYTHHHNQPHTHTHLDLPTYPFQHQHYWLESTQPGAGSG
SGSGSGRAGTAGGTAEVESRFWDAVARQDLETVATTLAVPPSAGLDTVVPALSAWHRH
QHDQARINTWTYQETWKPLTLPTTHQPHQTWLIAIPETQTHHPHITNILTNLHHHGIT
PIPLTLNHTHTNPQHLHHTRQQAQNHTTGPITGLLSLLALDETPHPHHPHTPTGTLLN
LTLTQTHTQTHPPTPLWYATTNATTTHPNDPLTHPTQAQTWGLARTTLLEHPTHTAGI
IDLPTTPTPHTLHHLTQTLTQPHHQTQLAIRTTGTHTRRLTPTTLTPTHQPPTPTPHG

TTLITGGTGALATHLTHHLTTHQPTQHLLLTSRTGPHTPHAQHLTTQLQQKGIHLTIT
TCDTSNPDQLQQLLNTIPPQHPLTTVIHTAGILDDATLTNLTPTQLNNVLRAKAHSAH
LLHQLTQHTPLNAFVLYSSAAATFGAPGQANYAAANAYLDALAHHRHTHHLPATSIAW
GTWQGNGLATGQVSEHLRRRGMFAMPPELAVTAVDGAIASGRPSLLVADIDWKKLGPV
LSSKSSVLLEDLPQAQGTEEARSTVEQTESTNLRQLLMGRSRSEQEEELLSLVRIHSA
AVLGRDDSEAIPPGRLFRDLGFDSLAAVELRNHLAAQTELALPTTLVFDYPSPTKLAQ
FLLSEIAEFQPDNSTPLPRPRAELDEPIAIVGMACRFPGGVTSADDFWDLISSEQDAI
GGFPTDRGWDLDTLYDPDPDHPGTCYTRNGGFLYDAGHFDAEFFGISPREALAMDPQQ
RLLLETAWETIEHAGINPHTLHGTPTGVFTGTNGQDHAAHIRQAPSGTEGFVLTGAAT
SIASGRISYILGLEGPAVTLDTACSSSLVALHLACQSLRSGECTMALAGGATVMTTPI
TFTEFARQRGLAPDGRCKAFSAAADGTGWGEGVGMLLVERLSDARRNGHRVLAVVRGS
AVNQDGASNGLTAPNGPSQQRVIRQALANADLTPADVDAVEAHGTGTTLGDPIEAQAI
LATYGQDRPGNGPLWLGSVKSNVGHTQAAAGVAGVIKMVMALRHRTLPPTLHADEPSP
HVDWSAGAVQLLTETVPWPGGEGRPRRAGVSSFGVSGTNAHVILEEAPADDVPGGPPA
DEDAGSGEEAAAGSPGVWPWLVSAKSQPALRAQAQALHAHLTDHPGLDLADVGYTLAH
ARAVFDHRATLIAADRDTFLQALQALAAGEPHPAVIHSSAPGGTGTGEAAGKTAFICS
GQGTQRPGMAHGLYHTHPVFAAALNDICTHLDPHLDHPLLPLLTQNDNDNEDAAAAL
LQQTPYAQPALFAFQVALHRLLTDGYHITPHYYAGHSLGEITAAHLAGILTLTDATTL
ITQRATLMQTMPPGTMTTLHTTPHHITHHLTAHENDLAIAAINTPTSLVISGTPHTVQ
HITTLCQQQGIKTKTLPTNHAFHSPHTNPILNQLHQHTQTLTYHPPHTPLITANTPPD
QLLTPHYWTQQARNTVDYATTTQTLHQHGVTTYIELGPDNTLTTLTHHNLPNTPTTTL
TLTHPHHHPQTHLLTNLAKTTTTWHPHHYTHHHNQPHTHTHLDLPTYPFQHHHYWLEL
PSAQTSPGQRRSRRSAPDTAESEFWDAVNEEDLQSLAETLDIDASALDTVVPALSAWH
RHQHDQARINTWTYQETWKPLTLPTTHQPHQTWLIAIPETQTHHPHITNILTNLHHHG
ITPIPLTVNHTHTNPQHLHHTLHHTRQQAQNHTTGPITGLLSLLALDETPHPHHPHTP
TGTLLNLTLPQTHTQTHPPTPLWYATTNATTTHPNDPLTHPTQAQTWGLARTTLLEHP
THTAGIIDLPTTPTPHTLHHLTQTLTQPHHQTQLAIRTTGTHTRRLTPTTLTPTHQPP
TPTPHGTTLITGGTGALATHLTHHLTTHQPTQHLLLTSRTGPHTPHAQHLTTQLQQKG
IHLTITTCDTSNPDQLQQLLNTIPPQHPLTTVIHTAGVNLFAPVSETDAESFSSVTAA
KATGAAILHELLLDHETLEHPILFSSGAGAWGSGNQCAYSAANAYLDALATHRQTHGL
PGASIAWGPWAGKGMSAGDAAHGYLEKRGILPMEPRMALAAFHRARAQRPNSNLIIAD
IDWERFVPAFTARRHSPLIEDIPEVRQAAQELEAAASTAKTTTAQPIATSLRERLARL
TSSKQNQVLLGLIRTGICTVLGLRNPEGIEDQRAFRDLGFDSLTSAQFSKELAKETGL
PLPPSLVFDYPTPQECAAHLRTQLVDLDDEEDAALSNALPQVAHRRTVEDEPIAIIGM
ACRFPGGVRSADDLWELLASGKDAIGVFPTDRGWDLDTLYDPDPDHPGTCYTRNGGFL
YGAGHFDAEFFGISPREALAMDPQQRLLLETAWETIEHAGINPHTLHGTPTGVFAGIN
AQDHAAHIRQSRDVETIEGYALTGSSGSVASGRVAYTLGLEGPAVSVDTACSSSLVAL

Query Match 62.5%; Score 273.6; DB 1; Length 64957;
Best Local Similarity 78.6%; Pred. No. 1e-41;
Matches 327; Conservative 0; Mismatches 89; Indels 0; Gaps 0;

```
Qy        23 CCGGTGCCGGCAACGTGTCAGCAGCCGGACTCGACCCCACCGAACACCCCCTACTCGGCG 82
             | || || || |||| | || || || | || || |||||||||| || |||| |
Db     47153 CGGGAGCAGGTGACGTCGCCGCTGCTGGCTTGGAGCCGGCCGAACACCCTCTGCTCGCGG 47094

Qy        83 CCACATTGGAACTGGCGACTGACGGTGGAGCGCTTCTTGCAGGGCGCTTGTCTTTGAGGT 142
             | ||| | |||| || | || || || | || ||| |||| ||| | |
Db     47093 CAACAGTCCAACTCGCAGACACGGACGGCTGCCTACTGACGGGTCGCCTGTCCTTGCGCT 47034

Qy       143 CGCATCCGTGGCTGGCTGACCATGCCGTCGGCGGCACGGTGCTGCTGTCGGGCGCCACCT 202
             |||||||||||||| || | || || || |||| |||||||||| | | |
Db     47033 CGCATCCGTGGCTGGGCGATTACGAGGTGGGGGGTGCGGTCCTGCTGTCGGGGTCGGCGT 46974

Qy       203 TCCTCGAACTCGCCCTTCATGCGGGCACATACGTGGGCTGCGACCGAGTGGATGAGCTGA 262
             || | || || || | || | ||| | |||||||||| |||| | || | ||| |
Db     46973 TCGTGGAGCTGGCGGTCCAGGTTGGCGAACGCGTGGGCTGCACCCGAATCGAGCAACTCA 46914

Qy       263 CGCTGCATGCGCCGCTGGTGGTTCCTGTGGATGGGGGTGTGAGTGTGCAGGTTGGGGTTG 322
             | |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db     46913 CTGTGCATGCGCCGCTGGTGGTTCCTGTGGGTGGGGGTGTGAGTGTGCAGGTTGGGGTTG 46854

Qy       323 CGGCTGCGGATGGGGAGGGGCGGCGTTTGGTGAGTGTGTATGCGCGGGGTGGGAGTGCTT 382
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     46853 CGGCTGCGGATGGGGAGGGGCGGCGTTTGGTGAGTGTGTATGCGCGGGGTGGGAGTGCTT 46794

Qy       383 GTGGTGGGGGTGGTGCGTCGGGTGGGGTGTGGACGTGTCATGCCTCGGGGGTGCTG 438
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     46793 GTGGTGGGGGTGGTGCGTCGGGTGGGGTGTGGACGTGTCATGCCTCGGGGGTGCTG 46738
```

RESULT 8
AF357202

```
LOCUS       AF357202              113193 bp    DNA     linear   BCT 17-JUL-2001
DEFINITION  Streptomyces nodosus amphotericin biosynthetic gene cluster,
            complete sequence.
ACCESSION   AF357202
VERSION     AF357202.1  GI:14794889
KEYWORDS    .
SOURCE      Streptomyces nodosus.
  ORGANISM  Streptomyces nodosus
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1  (bases 1 to 113193)
  AUTHORS   Caffrey,P., Lynch,S.V., Flood,E.M., Finnan,S.M. and Oliynyk,M.
  TITLE     The amphotericin biosynthetic gene cluster from Streptomyces
            nodosus
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 113193)
  AUTHORS   Caffrey,P., Lynch,S.V., Flood,E.M., Finnan,S.M. and Oliynyk,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-MAR-2001) Industrial Microbiology, University College
            Dublin, Belfield, Dublin, Ireland
FEATURES             Location/Qualifiers
     source          1. .113193
                     /organism="Streptomyces nodosus"
                     /db_xref="taxon:40318"
     gene            complement(4. .1824)
                     /gene="amphG"
     CDS             complement(4. .1824)
                     /gene="amphG"
                     /note="possible amphotericin export protein"
                     /codon_start=1
                     /transl_table=11
                     /product="AmphG"
                     /protein_id="AAK73498.1"
                     /db_xref="GI:14794890"
                     /translation="MAPDVPEEHEEERESEQPVRRLAALLRPHRRSVGLALTAGVVGI
                     LLNAFGPLLLGRVTDLIADGVLGHGGPAPGVDFGALGRLLMILLVLYVVASVFMLVQN
                     WLVASVVRLLIHDLRHRAQEKLARLPLRYFDRKPAGETLSRGTDDVDNLQQTLQQTLT
                     DLISSVFSLVIMLSLMLIISPSLAGVMLLSVPVSGLLAAWISKRAQPQYAAQWSASGK
                     LTAHVEEMCAGHALVKAFDRRAEAEQRFDERNEAVYRAGSGAQFASGAIEPVMMFVAN
                     LGYVAVAVVGAWKVVNGSLTLGDVQAFILYARQFSQPIVEIASVAGRLQSGVASAQRV
                     FTLLDAPEQEPEPDRPLAVERVEGRVEFQDVSFRYSPDTPLIEGLSLSVEPGSTVAVV
                     GPSGAGKTTVANLLMRFYEIDSGRILLDGTDTAAMNRDDLRSRFGLVLQDTWLFKGTI
                     AENIAYGSPGATRADIVEAARATYADRFIRTLSQGYDTVLDDESGGVSAGEKQLITVA
                     RAFLARPAVLVLDEATSSVDTRTELLIQRAMNTLRPGRTSFVIAHRLSTIRDADVIVV
                     MKSGRIVEQGTHDQLIDAQGAYARLHAARADAPAADVTVG"
     gene            complement(1805. .3628)
                     /gene="amphH"
     CDS             complement(1805. .3628)
                     /gene="amphH"
                     /note="possible amphotericin export protein"
                     /codon_start=1
                     /transl_table=11
                     /product="AmphH"
                     /protein_id="AAK73499.1"
                     /db_xref="GI:14794891"
                     /translation="MAPSVRWPTPFQRAAAGRRQPGCHSVLLRLMRTQLRPYAGSVVA
                     LVVLHLVQILGTLLLPTLGAALIDEGVVRHDSDRIGTIGATMAVVALVQIAAALGAAA
                     LGARTSTALGRDLRSAVFRRVLDFSAREIGRFGTPSLLTRSVNDVQQVQNLAQSGLGI
                     FVAAPLMCLGSVLLALRQDVTLALILVPMVLVVAVCFGLLLSRMAALYARLQQTLDRI
                     GRLLRERITGVRVVRSFARDAHEGERFTRTNEELLGLSLGVGRLIAVMLPSVLLLMNL
                     FTLGLLWVGARRIDSGSMQIGALSAFLSYLSLILMSVVMLAFVFLNVPRARVCAERIT
                     EVLQAETDVVPPASPRPMAGPAGQVELVGAEFRYPGAENAVLRDLSLTLRPGERVAVL
                     GSTGSGKTTLLHLILRLVDVTAGEVRIGGTDVRELDPSVMAAAVGYVPQRPYLFAGTV
                     ASNMRFGRPDATDEELWEVLRIAQHEGFVTRLGGLDTEIAQGGTTVSGGQRQRLAIAR
                     ALLRRPAIYLFDDSFSALDQSTEAALRKALVPYTEGATVITVAQRVASVRDADRIVLL
                     DQGGIAATGTHDALLRDSPTYREIALSQRTREETAHGAGRS"
     gene            3840. .4874
                     /gene="amphDIII"
     CDS             3840. .4874
                     /gene="amphDIII"
                     /note="mycosamine biosynthesis protein"
                     /codon_start=1
                     /transl_table=11
                     /product="AmphDIII"
                     /protein_id="AAK73500.1"
                     /db_xref="GI:14794892"
                     /translation="MPKRALITGITGQDGSYLAEHLLSQGYQVWGLIRGQANPRKFRV
                     SRLASELSFVDGDLMDQGSLVSAVDKVQPDEVYNLGAISFVPMSWQQAELVTEVNGMG
                     VLRVLEAIRMVSGLSMSRTAGTEGQIRFYQASSSEMFGKVAETPQRETTLFHPRSPYG
                     AAKAYGHFITRNYRESFGMYAVSGMLFNHESPRRGQEFVTRKISLAVARIKLGLQDKL
                     ALGNMDAVRDWGYAGDYVRAMHLMLQQDAPDDYVIGTGEMHTVRDAVRFAFEHVGLDW
                     KDYVVVDPDLVRPAEVEVLCADSSKAQAQLGWKPSVDFQELMRMMVDADLASVSRQNE
                     LDDLLLAHSW"
     gene            5042. .33574
                     /gene="amphI"
     CDS             5042. .33574
                     /gene="amphI"
                     /note="polyketide synthase multienzyme polypeptide housing
                     extension modules 9, 10, 11, 12, 13, and 14"
                     /codon_start=1
                     /transl_table=11
                     /product="AmphI"
                     /protein_id="AAK73501.1"
                     /db_xref="GI:14794893"
                     /translation="MDNEQKLRDYLKLATADLRRARRRVGELESASQEPIAIVGMTCR
                     YPGGVSSPEDLWRMVEAGENGVTPFPTDRGWDLEALASSPTSRGGFLHDAPEFDADFF
                     GISPREAVAMDPQQRVVLESAWEAFERAGINPTSVKGSRTGVFIGAMAQDYRVGPADG
                     AEGFQLTGNTGSVLSGRISYTFGTVGPAVTVDTACSSSLVAVHLATQALRAGECTLAL
                     AGGVTVMSGPGTFIEMGRQGGLSVDGRCRSFGDTADGTGWAEGVGILVLERLSDAIRN
                     GREILAVVRGTAVNQDGASNGLTAPNGPSQQAVIEQALVNARLSAGDIDVVEAHGTGT
                     TLGDPVEAQALLATYGQQRDEDKPLLLGSVKSNISHTQAAAGVAGVIKMVMAMRHGVL
                     PRTLLADEPTRHVDWSQGAVRVLTENTEWPATGAPRRAAVSSFGISGTNAHTIVEQAP
                     EPEPADPEDDAPSTPAAVTGVLPVLLSGRSPEVLRAQAAALLTTLGTGTPPAPADLAY
                     SLATTRTAFEHRAVLLASDLPELTGRLTAIAEGTDPAVLADTVTGTARTETRLAVLFT
                     GQGAQRLGAGRELAARFPVFAAALDAALDAFTPHLDVPLRKVLWGEDADRLDRTEYAQ
                     PALFAVEVALYRLLESFEVKPDHLAGHSVGEIAAAHVAGVFSLDDAATLVAARGRLMQ
                     ALPEGGAMVAVQASEDEVAPLLAGHEDLVSLAAVNGPSAVVLSGDETTVTELAARLAA
                     DGRKTSRLRVSHAFHSPLMAPMLDEFRNVVEGLTLHSPLLPVVSDVTGEPATVAQLTS
                     PDYWVDHVRQAVRFADGIDWLARHDVTAFLELGPDSVLSAMAQNCLDAAGSDALTVPA
                     LREGRPEDHTFTAALAALHTQGTALHWDACFTGTGARRTDLPTYTFQRRRYWPRAVQG
                     GAADLRSVGLGAAHHPLLSAAVSLADSEGALLTGRISLLSHPWLADHTVRGATLLPGT
                     AFLELAVRAGDEVGCDRVDELTLAAPLVLPEQGGVQVQLWIGNPDASGRRSVTVYGRP
                     DADEDAPWTSHATGVLSASRTTSDFDATVWPPADAETLPVDGLYERLAEGGFGYGPLF
                     QGLRAAWRRGDEVFAEVVLPESGHTDAESFGLHPALLDSALHAASFVDLDERAAGGLP
                     FSWEGVSLHASGATTLRVRLAPAAGDAVAIAVADDSGQLVLSADSLILRAVAAREIDA
                     AAALVRDALFRLDWVPVTAVAASGTAAALVGEDPFGLRALPQFGDLAVHPDLADLAAA
                     DGAVPDTVLLPLTGTGPDADPVTAAHRAATEALAAVRTWLEQDERFAASRLALVTRGA
                     TTGHDPAAAAVWGLVRSAQSENPGRFLLVDLDADQDTPALPAAALTSEEPQLAVRGEE
                     LRAARLVRRPASTAEAVPAFGGEGAVLVTGGTGGLGAVLARHLVAEHGVRELVLVSRR
                     GGAAAGAAELVAELAESGARATVVACDVTDRAAVAELVAAHPVSAVVHSAGVLDDGMV
                     GTLTPERLTTVLRPKVDAAWNLHEATRDLDLKAFVLFSSVAGVLGSPGQANYAAGNAF
                     LDALAAHRRAAGLPGLSLAWGPWEQTGGMTGGISEDDLRRMARAGTPALTVEQGLALL
                     DAALDGDDAALAPVRLDLSVLRAQGEVPPLLRSLIRGRSRRAAVAGSATAGGLAQRLA
                     RLDAESRDELVLDLVRGQVALVLGHATGAEIDAGRAFRELGFDSLTAVELRNRLNTVT
                     GLRLPATLVFDYPTVSHLASYVLDELLGTEVEAEVVQRGTAAVADDPIVIVGMACRYP
                     GGVTSPEDLWRLVTEGTDAVSGFPVNRGWDVENLYHPDPDHPGTAYTRSGGFLHEAGE
                     FDPGFFGMSPREALATDSQQRLLLEASWEAIERAGIDPVGLRGSATGVFAGVMYSDYS
                     AMLGSPEFEGFQGSGSSPSLASGRVSYTLGLEGPAVTVDTACSSSLVAMHWAMQALRS
                     GEISLALAGGVTVMSTPAVFVDFARQRGLSPDGRCKAFSDSADGVGWSEGVGMLVLER
                     QSDAIRNGHQILAVVRGSAVNQDGASNGLTAPNGPSQQRVIRQALASGGLSAGDVDVV
                     EAHGTGTTLGDPIEAQALLATYGRDRDPEQPLLLGSVKSNIGHTQAAAGVAGVIKMVM
                     SMRHGVLPRTLHVDAPSSHVDWTEGAVELLTEQTAWPETGRPRRAAVSSFGISGTNVH
                     TVLEQAPGTTVPAPAAPERTAGAVPLLLSGRTRDALRAQAARLLTHLQNHPEPSLADL
                     GHSLATTRSRFERRAAVIAQDREGLLASLGSLAAGRPDPAVVEGEAAGRARVAVMFSG
                     QGSQRAAMGRELYETQPRFAAAFDEVCAALDPLLDRPLREVVFAAEGSEEAALLDRTG
                     WTQPALFAVEVALYRLVESWGVRADFVTGHSIGEIAAAHIAGVFTLQDAARLVAARAT
                     LMEALPSGGAMVAVQATEEEVAPLLGEGLSVAAVNGPTSVVVSGDEDPAVELAAEFSG
                     RGRRTKRLRVSHAFHSPHMDAMLDAFRTVAETLSYAAPRIPLVSDLTGRRADDAEVRT
                     ADYWVRHVREAVRFADCVRTLRDAGATLFLELGPDGLLTAMAEDTLGDERYDHNTALV
                     PLLRADRPEESAAATAAARLQIHGVDLDWTAYLAGTGARRVDLPTYAFQHAHYWPQLP
                     SAAPSPAGDPADQKLWAAVERGDAAELAAVLGLDEDSLTPLDSLLPALSSWRRGNQEK
                     ALLDTLRYRVEWTRLSKPAAPVLDGTWLLVSSDATADDETELLDGLAEALGAHGARVR
                     RLVLDADCADRAVLGARLADTENADNTAQVLSVLPLDERPTDGPAGFTQGLALTIALV
                     QALADTGAHGRLWTATRGAVSTGPADPVTHPAQATAWGMGRGVALEHPRLWGGLVDLP
                     ADFDRGAGQRLAEVLAVKDAPDGEDQVALRATGVHGRRLVRHIVDELPSADQFTASGS
                     VLITGGTGGLGAETARWLARSGAAHLVLTSRRGPDAPGAAELRAELEQSGASVSIVAC
```

```
                DVADRDALAAVLDGLSADQPLTGVVHTAGVGHYGPLDALTPAEFAGLTAAKLAGAAHL
                DNLLGDRELDFFILFGSIAGVWGSGDQSAYGAANAYLDALALARRARGLAATSIAWGP
                WGGTGMAADDAVSGTLRRQGLGLLDPAPALTEMRRAVVRQDVTVTVADVDWTRYAPLF
                TSARPSALISDLPEVRALAAENTPADTGDASEIVQRVRSLSEPEQLRLLTDLVRTEAA
                TVLGHSSAGAVPEDRAFREIGFDSLTAVELRKHLGAATGLSLPSTMVFDYPTPLELAQ
                YLRAEMVGSVLEVAGPVATGGTDDEPIAIIGMSCRYPGGVSSPEQLWDLVLSGTDAIT
                DFPVNRGWNTAGLYDPDPDHPGTTYSTQGGFLHEADEFDPMFFGISPREALVMDPQQR
                LLLETTWEAFERAGLTPDTLRGSLTGTFIGSSYQEYGMGAGDGAEGHLVTGTSPSVLS
                GRLAYVFGLEGPAVTVDTACSSSLVALHLACQALRNGESNLAVAGGATVMTTPNAFVA
                FSRQRALAQDGRCKAFSESADGMTLAEGVGIVLVERLSDARRNGHPVLAVIRGSAINQ
                DGASNGLSAPNGPSQQRVIRQALANARVAPGEIDLLEAHGTGTPLGDPIEAQALFATY
                GRTRTPETALLLGSVKSNIGHSQSAAGVASIIKMVMALRHGVMPQTLHADEPSSHVDW
                SPGTVRLLGENTDWPQTGRPRRAAVSSFGISGTNAHVILEQETEAPAAEDEQLAPAPL
                PVAAGVVPWLLSARGAAALREQADRLLTHLVTADPAARPIDIGLSLATSRALFEHRAV
                VVPPAGTDPLEALRAVAADGPSGVVARGVADVAGRTVFVFPGQGSQWAGMGAQLLDES
                PVFAERIAECAAALAEFTDWNLIDVLRGAEGAPTLERVDVVQPASFAVMVSLAAVWRA
                QGVEPDAVVGHSQGEIAAAVVSGALSLRDGARVVTLRAQAIGRSLAGRGGMMSVALPV
                AEVEARLEAFEGRVSVAAENGPRSSVVAGEPEALDELHAQLTAEEIRARRVAVDYASH
                SPHVEDLHDEILELLAEVAPRTSEIPFFSTVTGDWLDTTVMDAGYWYRSLRGRVLFAD
                AVRDLIAADHRAFIEVSSHPVLAMSVQDMIDDAGVAGVASGTLRRDNGGLDRFLLSAA
                EVFVRGVQVDWAAVFEGTGASRVDLPTYAFQHENLWAMAAAPEAVTAADPEDAAFWTA
                VEDGDVSALTAALGTDEDSVAAVLPALSSWRRARKERSTVDSWRYRPTWKPVTKLPQR
                TLDGTWLLVSADGVDDTDVAEALETGGAEVRRLVLDESCTDRAVLRERLTDADGLTGI
                VSVLAGAERTGAVPGTGLVLGVALTVALVQALGDAGIDTPLWALTRGAVSTGRSDKVT
                APVQAQVTGIGWTAALECPGRWGGVVDLPETLDARAGQRLAAVLAGALGDDDQIALRS
                SGVFTRRIVRADAAPDGSARDWKPRGTTLVTGGSGTLAPHLARWLAEQGAEHLVLVSR
                RGPEAPGAAELRAELAERGTETTLAACDITDRDAVAALLESLKAEGRTVRTVVHTAAT
                IELHTLDATTLDDFDRVLAAKVTGAQILDELLDDEELDDFVLYSSTAGMWGSGAHAAY
                VAGNAYLAALAEHRRARGLTALSLSWGIWADDLQLGRVDPQMIRRSGLEFMDPQLALS
                GLKRALDDDEQVIAVADVDWETYHPVYTSARPTPLFDEVPEVQRLTAAAEQSAGDPAR
                GEFAAALLALPAAEQHRKLLETVRTEAASVLGLSSAEDLTDQRAFRDVGFDSLTAVGL
                RNRLASVTGLTLPSTMVFDYPNPAALAGFLHSELADVHSAGAVAVTAGAPVDDDPIAI
                VGMSCRYPGGITSAEQLWRVSLEEVDAVSVFPADRGWDAEALYDPDPDASGRTYSVQG
                GFLRDVADFDPGFFGISPREALSMDPQQRLLLETAWEVFENAGLDPVAQRGSRTGTFI
                GASYQDYGAAVPGSEGSEGHMITGSLPSVLSGRVSYLFGLEGPAVTLDTACSSSLVAI
                HLACQSLRNGESTLALAGGASIMSTPMSFIGFSRQRALAEDGRCKAYAEGADGMTLAE
                GVGLILLERLSDARRNGHEVLAVIRGSAVNQDGASNGLTAPNGPSQQRVIRQALANAG
                VEANDIDVLEGHGTGTALGDPIEAQALFATYGKDRDPERPVLLGSVKSNIGHTQMASG
                VASIIKLVHALREGVAPKSLHIDQPSTHVDWSSGTIQLLTERTEWPETGRPRRAAVSS
                FGLSGTNVHTVLEQAPAADAPAAEDTPAPRDALVPVLVSGRGEAALRAQAGALLDLLA
                ERPGIHPTDLAFSLATSRAALEHRAAVVADDHEALVRGLTALRDGLPGAGLVQGRTGR
                GRTAFLFTGQGSQRLGMGRELYERHPVFADALDAVLARIDGTTERPLRDVLFAAEGSQ
                DAALLHRTGYAQPALFALEVALFRLLESWGVTPDYLAGHSVGEIAAAHVAGVLDLDDA
                CTLVAARGRLMQALPEGGAMVALEAAEDEVLPHLEGLADQVSVAAVNGPRSVVVAGEE
                EPVLALAAHFAEQGRRTKRLRVSHAFHSPLMDPMLDDFAAVARALTYHAPSIPFVSNV
                TGTLAAPEQVCTADYWVSHVRSAVRFADGIGWLSTQGGVQTFLELGPDGVLSGMARES
                LTDASRTALLPTLRGDRPEEQALVTAVAAAHAHGFDVDWTAWFQGSGARRVALPTYAF
                QRERYWPDTTAAGITAPAPGSALDAEFWAAVEHADVASLTASLGLDDATVTAMVPALT
                AWRQRRGEQSALDSWRYRVTWKPRGGAPGAAPTGRWLVLVPAEHRDEATAAWAADVEA
                ALATATVRVEVTGTDRAALAARLTEAADGDTFQGVLSLLALAPGDAGHPGAPAALTLT
                ATALQALGDARIDAPLWNITRGAVAVGRSEQVTAPEQAAVWGLFRAAALELPARVGGS
                VDLPEDLDTQAARRLRGILAAADGEDAVAVRASGVFLRRLAHSPAADTVGSAFDPAAG
                TVLITGGTGGIGGHLARRLARDGAAHLLLTSRRGPDAPGAGELRAELEESGARVTIAA
                CDAADRDALAALLATVPEDAPLTAVFHTAGVVDDHVVDELTPESFATVLHAKTVAARH

  Query Match          35.2%;  Score 154;  DB 1;  Length 113193;
  Best Local Similarity  65.3%;  Pred. No. 1.4e-19;
  Matches  226;  Conservative    0;  Mismatches  120;  Indels    0;  Gaps    0;

Qy         23 CCGGTGCCGGCAACGTGTCAGCAGCCGGACTCGACCCCACCGAACACCCCCTACTCGGCG 82
              ||| | |||  ||||    || ||||| |||| | || |||| ||||| || || ||||
Db     109849 CCGCCGGCGGTGACGTCCGTGCCGCCGGTCTCGGCTCCGCCGAGCACCCGCTGCTGGGCG 109908

Qy         83 CCACATTGGAACTGGCGACTGACGGTGGAGCGCTTCTTGCAGGGCGCTTGTCTTTGAGGT 142
              || |  | ||||| ||  | | ||  || | |||  |  | || ||| ||||  |||| |
Db     109909 CCGCCGTCGAACTCGCCGCGGGCGAGGGGGTGCTGTTCACCGGCCGCCTGTCGCTGAGCT 109968

Qy        143 CGCATCCGTGGCTGGCTGACCATGCCGTCGGCGGCACGGTGCTGCTGTCGGGCGCCACCT 202
              |||| ||||||||||| ||||||| ||||    ||||| || |||||| ||||| |  ||
Db     109969 CGCACCCGTGGCTGGCCGACCATGTCGTCATGGGCACCGTCCTGCTGCCGGGCACGGCCC 110028

Qy        203 TCCTCGAACTCGCCCTTCATGCGGGCACATACGTGGGCTGCGACCGAGTGGATGAGCTGA 262
              | || ||  | ||  | |  |||||     | || |||||| ||| || || || || | |

Db     110029 TGCTGGAGATGGCGGTCCGCGCGGGTGACGAGGTCGGCTGCGGCCGGGTCGAGGAACTCA 110088

Qy        263 CGCTGCATGCGCCGCTGGTGGTTCCTGTGGATGGGGGTGTGAGTGTGCAGGTTGGGGTTG 322
              | ||    |||||||||||  | || | |   || |    |    |||||||| | || |
Db     110089 CCCTCGCCGCGCCGCTGGTCCTGCCCGAGCGGGGCGCGCTCCAGGTGCAGGTCGCCGTCG 110148

Qy        323 CGGCTGCGGATGGGGAGGGGCGGCGTTTGGTGAGTGTGTATGCGCG 368
              | ||  |||| |     || |||||  |||  | |||||  ||||
Db     110149 CCGCCCCGGACGACTCCGGCCGGCGCACGGTCGGGGTGTACTCGCG 110194


RESULT 9
AB070949
LOCUS       AB070949               78210 bp    DNA     linear   BCT 11-OCT-2001
DEFINITION  Streptomyces avermitilis polyene macrolide biosynthetic gene
            cluster.
ACCESSION   AB070949
VERSION     AB070949.1  GI:15824136
KEYWORDS    .
SOURCE      Streptomyces avermitilis DNA.
  ORGANISM  Streptomyces avermitilis
            Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
            Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1
  AUTHORS   Omura,S., Ikeda,H., Ishikawa,J., Hanamoto,A., Takahashi,C.,
            Shinose,M., Takahashi,Y., Horikawa,H., Nakazawa,H., Osonoe,T.,
            Kikuchi,H., Shiba,T., Sakaki,Y. and Hattori,M.
  TITLE     Genome sequence of an industrial microorganism Streptomyces
            avermitilis: deducing the ability of producing secondary
            metabolites
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 98 (21), 12215-12220 (2001)
  MEDLINE   21477403
REFERENCE   2  (bases 1 to 78210)
  AUTHORS.  Ikeda,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-SEP-2001) Haruo Ikeda, Kitasato University, Dept. of
            Microbial Chemistry, School of Pharmaceutical Sciences; 5-9-1
            Shirokane, Minato-ku, Tokyo 108-8641, Japan
            (E-mail:ikeda@mc.pharm.kitasato-u.ac.jp, Tel:+81-3-5791-6242,
            Fax:+81-3-3444-6197)
FEATURES             Location/Qualifiers
     source          1. .78210
                     /organism="Streptomyces avermitilis"
                     /db_xref="taxon:33903"
                     /note="Actinobacteridae"
     CDS             complement(1. .1254)
                     /codon_start=1
                     /transl_table=11
                     /product="putative lipoprotein"
                     /protein_id="BAB69301.1"
                     /db_xref="GI:15824137"
                     /translation="MGEALITLCTRSTRLAATAGVLGGVLAVTALGGTSNAATSGNGA
                     GNHPTSQARINQAGAQEASDARIEITPGQGTYSVGTNHPVSVTVTSGKLTKVTMTAVA
                     TGAEIPGTLSTDGTSWKPNGRLERATKYQIAAEAGDVKGRSTTGSATITTVSPANDFI
                     GHLSPEDGSTVGVAMPVTVNFDKAISDKAAVQSEIRVSSSSGQRVVGHWFNDHRVDFR
                     PENYWRPGSTVTVTLNRHGIQKTVTFKIGRSQISTVDAKTKQMTVVRDGKTIKTIPIS
                     SGSPEHPTYNGQMVISEKFRQIHMNGASVGLTEKNGKPSYDIKAVPHAMRLTDSGTFI
                     HGNYWGTDSVFGEANTSHGCVGLKDVRGDSHGRQPAAWFFDHSIVGDVVIVKNSEEKT
                     MLAPDNGLSDWNMSWSEWVAGSPAT"
     CDS             complement(1256. .2290)
                     /codon_start=1
                     /transl_table=11
                     /product="putative transposase"
                     /protein_id="BAB69302.1"
                     /db_xref="GI:15824138"
                     /translation="MPVEFLSHDQVSRYGRFASAPTPGELEQFFRLDAKALESARSKR
                     APANRLGWAVQWGCVRMLGVFPTEDLSVVPEVVVRFAAEQLGVDPGEFAAYGARRQNR
                     YEHAWEIRDAYAYRESPAAEAQVREFLATRVWASWEGPRALFDRAVGWLVDNRVLLPG
                     ITTLTRLVAEVRAQENAALYRTLDEAVPEKLRQSMRDLLKVPEGRRVSELERLRRSPM
                     GVSGSADRLAFSERIRMEAVAMFTAGPGSTDIAKELRVSVQSVQRWRSGLAGDRPGRG
                     AVPQPVSVELSETLFAALDEERAKGFAPSWTIRYLHHDFPIFGQRYAYKWSQGCLHSF
                     GRSAQGYRDQ"
     gene            2935. .21081
```

```
                     /gene="pteA1"
CDS                  2935. .21081
                     /gene="pteA1"
                     /codon_start=1
                     /transl_table=11
                     /product="modular polyketide synthase"
                     /protein_id="BAB69303.1"
                     /db_xref="GI:15824139"
                     /translation="MVESELNATGNRDCSASADSADGIAVIGLACRLPGAADPAAFWR
                     LLSEGADAITDVPPDRWDGAVVADASAPDRTGIRRGGFLDSIGDFDAGFFGLSPKEAA
                     AMDPQQRLVLELAWEALEDAHVRAATLRSTRTGVFVGAMWDDYATLHHRSGLTAISPH
                     TVTGLHRSIIANRVSYFLGLNGPSLTVDSGQSSSLVSVHLACESLRKGESTIALAGGV
                     NLNIVPESALGAAKFGGLSPDGRCFTFDARANGYVRGEGGGIVVLKPLARAVADGDPV
                     YCVIRGSAVNNDGGGDGLTVPLQSGQEQVLRLAYERAGVNPADVGYVELHGTGTKVGD
                     PIEAAALGAVLGTGREPGRPLRVGSAKTNVGHLEGAAGITGLLKAVLSLNHRELPASL
                     NFGTPNPAIPLDRLNLRVQTEHSEWAAEDGRPLLAGVSSFGMGGTNCHIVLASGAAVE
                     AGEGATAGGGAAPTFWPLSAKTASGLRAQAAALLAHTEAPLGLALPDVGWSLATGRTA
                     FEHRAVVVGEVRDDFLRALRELSTGGIDAALTTGRAGPRGELAFLFSGQGSQRAGMGR
                     ELAESVPAFASALDEVCAHLDPLLPRPLREVMFAPVGTGEAAELDQTLYTQTSLFAVE
                     VALFRLLESWDITPDVLMGHSIGELAAAHVAGVLSLADACALVAARGRLMQALPAGGA
                     MIAVQAAEDEVRARIGERTDRVSIAALNGPDSVVVSGDEDLATEIADAFAALGRKTSR
                     LRVSHAFHSPHMEPMLAEFRNVAEGLTFHAPRIPIVSNVTGRLSEEPSGYEGCTAAYW
                     VRHVREAVRFADGVARLDEQGVRTYLELGPGGVLASMARAGADSESLFVPALRTRRPE
                     AQALLAAVASLHVHGVEPDWDALFGGGRGARRKVALPTYAFERQRYWLDGDTTSPVAT
                     TGWAPVAPQPDAIAETDSAPVGALGRRLAGLAESERDEAVLDMVRAHIAAVLGHAEAR
                     QVETEWTFKDLGFDSLSSVELRNQLSEATGLRLPSGLLFDHPTPAALARHLAAQSLGG
                     GDAAAAAPATVADPHEPIAIVAMSCRLPGGVSSPEDLWRLLASGGDGISGFPTDRGWD
                     VETLYDPEPGTPGKSSTRHGGFLHEAAEFDAGFFGISPREAAAIDPQQRLVLETAWEA
                     FERAGIDPATLRGSRAGVFIGATAQDYGPRLHEPADGVEGYLLTGTTTSVASGRVAYS
                     FGLEGPAVTVDTACSSSLVALHLAVQSLRSGECTMALAGGVTVMSTPGMFVEFSRQRG
                     LSADGRCKAFAAGADGTGWAEGVGMLLVERLSDARRNGHQVLAVVRGSAVNQDGASNG
                     LTAPNGPSQQRVVRQALAAAGLAPADVDAVEAHGTGTLLGDPIEAAALIAAYGQGRDG
                     DRPLWLGSLKSNIGHTQAAAGVAGVIKMVMAMRHGVLPRTLHVDEPTPHVDWSASGVR
                     LLTEAVEWPARDLPRRAAVSSFGISGTNAHVIVEQAATSEEPARGTAVAGPVPWVVSA
                     KSDVALREQAGRLLSFLDAEGVGGAGVRPEDVALSLATSRAVLERRAAVVGGDLAELR
                     QGLEALASGASGAAPVGGAVGGRVGFLFSGQGSQRIGMGRELYAAYPVFAAAYDEVCA
                     FLDAPVDVDSEELNQTGTTQPALFAVEVALFRLLESWGIRPDYVAGHSVGEIAAAHVA
                     GVLSLEDAAKLVSARAALMQALPDGGAMVAVQATEDEVLRHLTDGVGIAAVNGPRSVV
                     VSGAEDAVTAVGEIFAGRGRKTSRLKVSHAFHSPLMDPMLEEFAAVVRGLTFSESRIP
                     VVSNLTGRLAEVYTPEYWVRHVREAVRFADGIETLGELGVTTFVEVGPGGVLSALAQG
                     CFDDGTDIVTVPVLRADRPEPYALVTAIGRLHTHGISPDWQVLFPGARRVDLPTYAFQ
                     RERYWLDAPRAATAGAPDAMDAEFWDSVESEDRASLGALLGLEPTELDVVASKLSAWR
                     RKRREQSTTDGWRYRITWQPLDDMVRAAPSGTWLYVVSEETAWTEAIRAGLTELGVAL
                     VPFAVGEDTDRDALARAFADAGQEQADGILFAAAQGEAGANALQRLVLLVQALGDAGI
                     EAPLWCLTSGAVSTGTADLLTDPVSAQLWGLGRIVALEQPQRWGGLVDLPAELDSRAL
                     ERLAGLLAQADEDQLAVRASGVYGRRLVRAPLSAAPADAPWSPRGTVLVTGGTGALGG
                     QVARWLAGAGAEHLVLTSRRGPDAPGAAELKAELEELGAQVTVVACDAADRDALAELF
                     GRYIVNAVVHTAGVLDDGLVESLTPERLNHVLRPKVDAALHLHELTRDRQDLDAFVLF
                     SSMTGVWGNGGQGAYGAANAFLDALAEHRRAQGVPALAVAWGSWADGGMADGAAGDHL
                     RRRGVRAIAALPAISVLHGALTHGETSMTVADVDWERFVPAFVGTRPCPLLLGVPEAR
                     EALEARARAAQSPEVPVSALVQRLLGATPGEQGRILLDLVREQAATVLGHTGKGAFEA
                     DRTFRETGFDSLTAVELRHRLNTATGLRLPTTLVFDHPTPTALARHLGEELLGRHGLE
                     TAAAQATAAAAADEPLAIVGMSCRLPGGVQSPEDLWELLSSGRDAMSGFPTDRGWDID
                     ALYDPDPDRPGKTYSREGGFIQGADRFDAGLFGISPREALAMDPQQRLLLETAWEAFE
                     RAGIALASVRASRTGVFIGTNGQDYANGLRNAPEEIEGYALTGKAASVVSGRISYTFG
                     LEGPAVTVDTACSSSLVALHLAAQALRSGECSMALVGGVTVMTTPDLFVEFSRQRGLS
                     PDGRCKAFAAGADGTGWGEGVGLLLVERLSDARRNGHQVLALVRGSAVNQDGASNGLT
                     APNGPSQQRVIRQALASAGLTPADVDAVEAHGTGTRLGDPIEAQALLATYGQEHSDDR
                     PVWLGSVKSNIGHTQAAAGVAGVIKMVLAMRHGVLPQTLHVDEPTAHVDWSAGAVRLL
                     TEPVQWPGTTGPRRAAVSSFGIGGTNAHTIIEEAPAAAAAEPAREHRPVPVPWVLSAK
                     SEVALRAQAERLLAFATDDVSPVDAGFSSATTRSALEHRAAVIGSDPAELRAGLEALA
                     AGEPAANVVAGRAHSADKVGFLFSGQGSQRLGMGRELYAAYPVFAAAYDEVCALLDAT
                     VDVDAETLHQTGSTQPALFAVEVALFRLLESWGIRPDYVAGHSVGEIAAAHVAGVLSL
                     EDAAKLVSARAALMKALPAGGVMVAVQAAEDEMLPHLIDRAGIAAINGPRSVVISGAE
                     DAVTAIAEVFAAQGRKTSRLKVSHAFHSPLMDPMLEEFAEVVRGLTFSEPWIPVVSNL
                     TGRLTESYTPEYWVRHVREAVRFADGVLTLHELGVTTFVEIGPGGVLSALVQGCLDDG
                     TDIVTIPALRADRPEPQALTTAYAQLHVSGAEVDWHAFFPGARRVELPTYAFQRERYW
                     LDAPASAGDMRAVGQGEAGHPLLGAAVPLADSVGHLLTGRLSKHTHPWLVDHAVSGVA
                     LLPGTAFVELAITAADAVGCDLLEELTLETPLLVPERGGVALQVHVGADDGSGRRPLT
                     VHSQIDDGDAYAHEDDSERAWVRHASGVLTVATGGPVDGSLTAWPPVGAEPVDVDGFY
                     DRLAGMALDYGPVFQGLRAAWRSGDDFFAEVELPGAERTDAGAYGLHPALFDAALHTV
                     WLGAVEPEAGTGNGLLPFAWSGVRLAAAGASDLRVKVSRAGTGTSTVSLVLADGTGGP
                     VAEVDSLTLRPVPADQLRAGSGSDDAVFGLKWTPVSLPSAPDGMRIETYEDLAALRAT
                     DPPADTEADAAAAADAVVVPCPTGTGADLATRVREVTHAVLELVQWWLAEERPGRLVL
                     VTRGGDLAQSAVCGLVRSAQTENPDRIVLVETSSDAHAGAGSGVDDADLARVLPGLIA
                     SGEPQAAVHADDEVRVPRLVRVAVPDAEPELPGLGSGTVLLTGASGGLGGLFARHLVA
                     EHGVRSLLLVSRRGGDAPGAAELTADLTAQGADVTWAACDVADRAAVRALLAGPGQSL
                     SAIIHTAGVLDDGIIGSLTPERLDAVFRPKVDAALNLHDLAGELSGDLSAFVLFSSVA
                     GTLGTPGQANYAAANTFLDALAEHRRAAGQRATSLAWGLWAQDSESAMTGSLDHTDLT
                     RIQRMGLAAIPSSDGLRLFDAALALDRAVVAPLRLDLGGLQGTVPPVLRGLRGTVRGT
                     ARRTAQTAPKGSLGQRLAGLAEAEREQVVLDLVRTEVAAVLGHAGAQSVGADDSFKDI
                     GFDSLTAVELRNRLNSAVGMRLPATLIFDYPNPLTLAHFLTGEAAGGEAATAVPTTLS
                     TATGTADEPIAIVGMACRYPGGVQSPEDLWQLVFSGRDAVSGFPEGRGWDVEQLYDPN
                     PDQWGTSYTREGGFLHDAADFDAEFFGISPREALAMDPQQRLLLETSWEAFERAGIDP
                     ATVRGSRTGVFAGVMYHDYGGRVHTSPPGLEGYLVNGSAGSIASGRVSYTFGLEGPAV
                     TVDTACSSSLVALHLAAQALRSGECSMALVGGVTVMAGPSVFVEFSRQRGLSADGRCK
                     AFGAGADGTGWAEGVGMLLVERLSDARRNGHKVLAVVRGTAVNQDGASNGLTAPNGPS
                     QQRVIRQALANAGLSSDEVDAVEAHGTGTRLGDPIEAQALIATYGQERSEDRPLWLGS
                     LKSNIGHSQAAAGVGGVIKMVMAMQHGVLPQTLHVDEPTPHVDWSAGAVQLLTEAVEW
                     PASDGPRRAAVSSFGVSGTNAHVIVEQAATSEVTARGTAVAGPVPWVLSAKSDAALRE
                     QAGRLLSFLDAEGVGGAGVRPEDVAYSLATSRAVLECRAAVVGEELGEFRQGLEGIVS
                     GAAPVGGAVGGRVGFLFSGQGSQRIGMGRELYAAYPVFAAAYDEVCALLDATVDVDAE
                     TLHQTGSTQPALFAVEVALFRLLESWGIRPDYVAGHSVGEIAAAHVAGVLSLEDAAKL
                     VSARAALMQALPAGGAMVAIQATEAEVLPHLTHGVGLAAINGPRSVVLSGSEDAVLDI
                     AEVFAQQGRKTSRLKVSHAFHSPLMDPMLEEFAEVVRGLTFSEPRIPVVSNLTGRLAE
                     VYIPEYWVRHVREAVRFADGLHTLHELGVTMFVEIGPGGVLSALVQGCLDDGTHIVTV
                     PALRADRPEPYALVTAVGQLHTHGISPEWQALFPGARRVDLPTYAFQRERYWLDAPQA
                     SGDVRAAGLGAADHPLLGAAIATADSDGAVLTGLLSLGTHPWLADHAVNGIVLLPGTA
                     FVELAIRAGDETGCGWIEDLTLELPLVVPERGGVTLQVAVGPEDESGRRQLSVHSRVD
                     DRSWLRHATGVLSSAEAPSAAKLGAWPPAGAEAVDLTAFYEGMAAAGLDYGPVFQGLR
                     SVWRDGDDVFAEVELPDESAAEAAAFALHPALLDAALHALAAGGLVSLDDSPALPFAW
                     SGVVVQAAGAAMVRVKLSRTATGSVTLAVADGAGDPVASVESLSLRAASAEQLPGAVG
                     GDQLFGLEWTRFTLPSATDALMIETVADFDALRESEGTPDVVVVPVVPCPAGSGRETA
                     DRVHSTTSEVLALVQWWLAEERAGRLALVTRPGDLAHAAVWGLVRSAQSENPDRIVLV
                     EAEETDEAVRVLPAALASGEPQFAVRDGEVFVPRLVKATRSATGTATGTPDFGVGPVL
                     LTGASGALGGLVARHLVAEHGVRSLLLLSRRGAEAPGAVELEAELAAWGAEVRWAACD
                     VADREAVSGMLNGLGEQSLSAVVHTAGVLGDGIVASLTPERMREVFRPKVDAVLNLHE
                     CTRDMGLAAFVVFSSVAGMVGSAGQASYAAANSFLDAFSAHRRREGLPAISLAWGVWE
                     QSGAMTDGLVEADRARMARSGVLPLPPEEGLRLFDAALASDEAVLAPVRIDTGALRAG
                     EAPPVLRALVPAAAARRTARAAAAPSASSLAERLEEMPEAEREKTVLDLVRAEVAAGL
                     GHASDRTVRPEHAFQDLGFDSLTAVELRNRLNKATGLRLPATLVFDHPTPALLARHVF
                     MAAAGAAEPALVDSLLADLDRVEQELVTTLAASEARDRILSKLQGLLAIAGEFGKAPA
                     QEGPGTGGSDLESATDDEVFDLLGKEFGIS"
gene                 21273. .31847
                     /gene="pteA2"
CDS                  21273. .31847
                     /gene="pteA2"
                     /codon_start=1
                     /transl_table=11
                     /product="modular polyketide synthase"
                     /protein_id="BAB69304.1"
                     /db_xref="GI:15824140"
                     /translation="MSCRYPGGVDSPEDLWELVAGGRDAVSEFPADRGWDLDALYDPD
                     PESRGTSYAREGAFLRDLARFDAGLFGISPREALAMDPQQRLLLETSWELFERAGIDA
                     ASLAGSRTGVFAGVMNHEFLAALQNSPQDLEGYLGTGTSGSVASGRVAYTFGLEGPAV
                     TVDTACSSSLVALHLAVQSLRNGECSLAVAGGVTAMVGPATFVGFSRQRGLAPDGRCK
                     PFAAGADGTGWGEGAGLLLVERLSDARRNGHEVLAVVRGSAVNQDGASNGLTAPNGPS
                     QQRVIRQALAAAQLTTAQVDAVEAHGTGTRLGDPIEAQALLATYGQGRPDDKPLWLGS
                     IKSNIGHTQAAAGVAGIIKMVMAMRHGELPRTLHVDAPSVHVDWTAGAVRLLTERTPW
                     SASDQPRRAGISSFGISGTNAHTIIEEAPSAEAFAQDTGAPEGPEEGAESAEPVASEA
                     ARPAAGGPVPWVISGRTEEALRAQAERLLSRTDAEVPPLDVAFSLATTRTALEHRAVL
                     QGETPAELAEGLRAVASGSPARGVARGLSGTGGRVGFLFSGQGSQRLGMGRELYASYP
                     AFAAAYDEVCAQLDAVVDVDSEELHRTGAAQPALFAVEVALFRLLESWGIRADYVAGH
                     SVGELAAAHVAGVLSLDDAAKLVSARARLMQALPAGGAMVAVQATEDEVLPHLTDAVG
                     IAAINGPRSVVVSGAEDAVTAIAEAFRQQGRKSSRLKVSHAFHSPLMDPILEEFAEVV
                     RSLTFHRPQIPVVSNLTGRLAEPYTPQYWVRHVREAVRFADGVRTLNDLGVTTFVEIG
                     PGGVLSALAEGCLDDDVVTVPALRADRPERVALVAAVSELHAHGVSPDWHAFFPGARR
                     VDLPTYAFQYERYWLDASPAAPGGVRAAGLGSADHPLLGAAVSLAGGDERLLTGRLSL
                     RTHPWLSDHAVLGTALLPGTAFVELAVRAADEAGCALLEDLTLEAPLVVPEQGGVAVQ
                     VWVGAVDGTGRRPLTVHSRPEDDSDLPWVRHATGFVTEGSEEAGAAAADRTLTAWPPT

 Query Match          34.0%;  Score 148.8;  DB 1;  Length 78210;
 Best Local Similarity  62.8%;  Pred. No. 1.4e-18;
 Matches  231;  Conservative    0;  Mismatches  137;  Indels    0;  Gaps    0;

Qy          1 TACTGGCTCGAAAGCACACAGCCCGGTGCCGGCAACGTGTCAGCAGCCGGACTCGACCCC 60
```

```
                ||||||||||| | |||  ||| | ||||  |||   || ||||| || | | |
Db      23877 TACTGGCTCGACGCCTCACCAGCCGCTCCCGGTGGCGTCCGCGCCGCCGGCCTTGGCTCG 23936

Qy         61 ACCGAACACCCCCTACTCGGCGCCACATTGGAACTGGCGACTGACGGTGGAGCGCTTCTT 120
               |||| ||||| || || |||||| |  |   |||||   | ||  |   ||| ||
Db      23937 GCCGACCACCCGCTGCTGGGCGCCGCCGTATCCCTGGCCGGCGGCGACGAGCGGCTGCTG 23996

Qy        121 GCAGGGCGCTTGTCTTTGAGGTCGCATCCGTGGCTGGCTGACCATGCCGTCGGCGGCACG 180
               | || ||  |||| || |  |||| |||||||||   ||||| |||||   ||| ||
Db      23997 ACCGGCCGGCTGTCCCTGCGCACGCACCCGTGGCTGAGCGACCACGCCGTGCTCGGAACC 24056

Qy        181 GTGCTGCTGTCGGGCGCCACCTTCCTCGAACTCGCCCTTCATGCGGGCACATACGTGGGC 240
               | ||||||| ||||| || | ||| ||||||| ||| | |  || | |    | |  |||
Db      24057 GCGCTGCTGCCGGGCACCGCGTTCGTCGAACTGGCCGTGCGCGCCGCCGACGAAGCCGGC 24116

Qy        241 TGCGACCGAGTGGATGAGCTGACGCTGCATGCGCCGCTGGTGGTTCCTGTGGATGGGGGT 300
               ||||    | |||| || || || ||  | ||||| |||||||| || | | | || ||
Db      24117 TGCGCATTACTGGAGGACCTCACCCTCGAAGCGCCCCTGGTGGTGCCCGAGCAGGGCGGC 24176

Qy        301 GTGAGTGTGCAGGTTGGGGTTGCGGCTGCGGATGGGGAGGGGCGGCGTTTGGTGAGTGTG 360
               ||    || |||||  |||| |  || |  || ||    || ||  |   | |||  |||
Db      24177 GTCGCGGTTCAGGTGTGGGTCGGCGCGGTCGACGGCACCGGCCGCAGGCCGCTGACCGTG 24236

Qy        361 TATGCGCG 368
               |  ||||
Db      24237 CACTCGCG 24244


RESULT 10
AB070940/c
LOCUS       AB070940              104326 bp    DNA     linear   BCT 22-JAN-2002
DEFINITION  Streptomyces avermitilis oligomycin biosynthetic gene cluster.
ACCESSION   AB070940
VERSION     AB070940.1  GI:15823967
KEYWORDS    .
SOURCE      Streptomyces avermitilis DNA.
  ORGANISM  Streptomyces avermitilis
            Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
            Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1
  AUTHORS   Omura,S., Ikeda,H., Ishikawa,J., Hanamoto,A., Takahashi,C.,
            Shinose,M., Takahashi,Y., Horikawa,H., Nakazawa,H., Osonoe,T.,
            Kikuchi,H., Shiba,T., Sakaki,Y. and Hattori,M.
  TITLE     Genome sequence of an industrial microorganism Streptomyces
            avermitilis: deducing the ability of producing secondary
            metabolites
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 98 (21), 12215-12220 (2001)
  MEDLINE   21477403
REFERENCE   2  (bases 1 to 104326)
  AUTHORS   Ikeda,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-SEP-2001) Haruo Ikeda, Kitasato University, Dept. of
            Microbial Chemistry, School of Pharmaceutical Sciences; 5-9-1
            Shirokane, Minato-ku, Tokyo 108-8641, Japan
            (E-mail:ikeda@mc.pharm.kitasato-u.ac.jp, Tel:+81-3-5791-6242,
            Fax:+81-3-3444-6197)
FEATURES             Location/Qualifiers
     source          1. .104326
                     /organism="Streptomyces avermitilis"
                     /db_xref="taxon:33903"
     gene            complement(1. .1134)
                     /gene="mcmA"
     CDS             complement(1. .1134)
                     /gene="mcmA"
                     /note="methylmalonyl-CoA mutase"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="BAB69185.1"
                     /db_xref="GI:15823968"
                     /translation="MTGYSTAADANARFRQLVGSGTRAVPVVFDLPTRAGLDSDAPAA
                     RGLVGRRGVAVDSIDDMRVLFGGLRLDRVVPSLAVDAPAAPLLVLYQLVAEEHGAPAQ
                     RLAGAVRGDVFKEFLARGPGVFPLRPSLRLAMNVFAYCQAELPRWQCLSVCGHHLAGT
                     GADPALEVAFTVAGGIEYLRAAVTAGLDVRTVAPRLTLCLVGGTTGRENRAKLRATRR
                     VWTRTANERFGSGRTAPRLYVRTPRDGLATLACTEDIEAGATALVDRIERLGGTLTVL
                     EHGLAGQLLGRPHGDRPATPRPHPPHQDSLDLQARQTERLAKLRAWRVQPAVDEALLR
                     LRKAAQGDDNVLYPMREALAARATVGEVCDALRSTWGTGPSAT"
     gene            complement(1131. .1886)
                     /gene="pptA"
     CDS             complement(1131. .1886)
                     /gene="pptA"
                     /codon_start=1
                     /transl_table=11
                     /product="phosphopantheteinyl transferase"
                     /protein_id="BAB69186.1"
                     /db_xref="GI:15823969"
                     /translation="MRAHEVFADVPVSSLHPREAALVQDAVESRRREFATGRHCARRA
                     LASLGVPPGPLLATPRGAPRWPAGVCGAITHCAGYRAAAVARAPGIRAVGIDAEPDRP
                     LPEGVLEAVALPAEQFQVKEHTARHPGVVWGKLLPSAKESVYKAWFPATGRPLGFEEA
                     LIDLDPVSSTFRVRVLPTALRAAEPELRDLHGRWLVRHGIIVTAVVVDGPGRDTPHAP
                     SPYDRTKHSAPTGRSADRPGGRPTDRATDRPTA"
     gene            complement(1910. .2875)
                     /gene="simX4 homolog"
     CDS             complement(1910. .2875)
                     /gene="simX4 homolog"
                     /codon_start=1
                     /transl_table=11
                     /product="SimX4 homolog"
                     /protein_id="BAB69187.1"
                     /db_xref="GI:15823970"
                     /translation="MSQDRHDGSTAALGGPSHGGRLLAVSDIHVRHRQNRDIVQDLHP
                     GCAEDWLIVAGDVAESVADITWALSLLRDRFAKVIWAPGNHELWTAPGDPDELRGERK
                     YRHLVDLCRALGVVTPEDPYPVWHGAGGPVAVAPLFLLYDYSFLPPGAWSKSHALRLA
                     DDAGVICTDEFLLHPDPYPSREAWCRARVRATRARLDALPDDLPTVLVNHFPLVREPT
                     RVLRHPEFALWCGTTATADWHLRYRAAAVVHGHLHIPRTIRCDDVPHHEVSLGYPREW
                     QRYGRRPDWPVQILTPAPAPAPAPVPVPVPAPPGHRTASPADRRP"
     gene            3213. .3971
                     /gene="olmC"
     CDS             3213. .3971
                     /gene="olmC"
                     /codon_start=1
                     /transl_table=11
                     /product="thioesterase"
                     /protein_id="BAB69188.1"
                     /db_xref="GI:15823971"
                     /translation="MSANRVDGNLWVRQFHPAPTAATKLVCLPHAGGSASYFLPVSKA
                     LAPGVDVLAIQYPGRQDRRHEKCIDNIAELADAVTAVLEPLADRPLVLFGHSMGATLA
                     FEVALRLERKGIVPLALFASGRRAPSRHRDESVHLRGDDEIIAELKTLSGTRSEVLGD
                     EEILRMVLPAIRSDYKAAETYRYAGGARLTAPIHAHFGVEDPRVSLDEAQAWADHTTG
                     HFELRSHPGGHFYLNDQAPRVIEEIARVSSSVAS"
     gene            4643. .7438
                     /gene="olmRI"
     CDS             4643. .7438
                     /gene="olmRI"
                     /codon_start=1
                     /transl_table=11
                     /product="luxR-family transcriptional regulatory protein"
                     /protein_id="BAB69189.1"
                     /db_xref="GI:15823972"
                     /translation="MLIGRGQEIDRLRAAYSSVRSTGRSRAVLVESAAGCGKSEILED
                     LVRHVESLGLIALRATAIAEERNTPLSVLRSLVNGAPFPDDLVRRFHTMSDAAARMIS
                     TEGRDKARRDAAWYELMRTFCTEVRALAQRAPLVVAVDDIHHGDADSLRQLLHLARHC
                     RNVPVLLTFTQSLAGGPADRTFGAELLRQPNLEHIRLPRFCRQDIADMLAEQGRADAG
                     QGTVERYFAVSGGNPLLLRALLEDARLEPDDTVDAQDVGPRAGQAYGRAVLACLDRGG
                     EKAHEVAAGLAVLGGAFTHELLSRLLDISAAALTENLRALRAAGIVVGTSFQHPAAKA
                     AVLDSLEPSHRLRLHQRSAALLHASGAPAALAAEHLHEAGQADGAWAVGVLQQAAEQA
                     LLADEARRAVSYLELAHAASADPRGRTEIKIRLGSVTRRVSAADAERHVDDALRGFGA
                     GLLTPSDTGGLAALLLAHGRLEQARSVLEQGRRPGDGDADRTAARRQLAEIRDSVLNR
                     GAWCAGTRAAWHENPAGRPSTGRAGTGLRAWLSGGSHAALLAQRIAGTEGLLEITHLTD
                     ATFDLVLNAVWTLLALGAVERALHWCDTFLREAKRRDAPGWEAVFAAVRAEGALRGGN
                     LAEAERDARRCFAVLPERHRSALEGSAVSYLVTVQIARGKYEDAARQLSRPMADSLPD
                     TIHWLDYLRARGLYYMATRQYHSALRDFHEVGRLAERWGTDWPALLPWRTDTAEALLR
                     LGEREQGLRLATEQLAMVQDGAPRIKGISLRLQGLASEPAARLGLLNRAVVELSRADD
                     RLELSRTLFDLADAHRETRQTRQAAMVLRRARQLAKSCGAQPPRDRIGSPDAAGASGE
                     LFGEGPERNADWGHLSESERRVAGLAASGYTNREISVRLYITMSTVEQHLTRVYRKLK
                     INGREELLAEARAGVQV"
     gene            7658. .10486
```

```
                     /gene="olmRII"
CDS                  7658. .10486
                     /gene="olmRII"
                     /codon_start=1
                     /transl_table=11
                     /product="luxR-family transcriptional regulatory protein"
                     /protein_id="BAB69190.1"
                     /db_xref="GI:15823973"
                     /translation="MLASAGQGEGGVVLVNAGPGVGRTALLDAFLRKATESGARVCAA
                     AGSSAETEAEFGVVGQLFPPTGPLGAGFRLARAIGSRDAGPGLDADRLFDALCQELRR
                     RPTVIAVDDAQLADAASLRFLRYLVRRLRTAPVMMVLTEPTGPDALPPSFQAELLRHP
                     QCRRLGLLPLSRDGVTQVVSAQVGSDRVPRLAADFHAISGGNPLLVHALLADHRAAHR
                     LSPVPSVPTVPSVSSASSAAASASSAASAASAVSAAPAASAAFAKAALAWAYRDEPVLF
                     DVACGFAVLGESATPARVACLVDRGADAVDQVMTALDTAGLLAGGTFRSALVGKALLE
                     HVDVETRAELRRRAARLLHGDGAPATVVARHLLAAPVAEPWAGRTLLAAAEKAMQEGR
                     EELSLDCLRLAGQSATTEQERAAVVMARVKIGWEIDPRLVSPWLDELVAALRRGDLRS
                     GDAAWIVKHLAWQDRFPEAVESLSTLAAAGDAGEPGVEAELHFVRHWLRHSCPPLEAS
                     VPGPSNPTRPRTPRLRRVSPASYAGEVLGKVLTEGPDDHTVAMAEEILRGCRYSETVV
                     EALEAALLALVYAGRPDRAAFWCDGLLQQAGTRSTGTSVAILSGIRAEIALRQGALPA
                     VEQYARSALSAISLPGWGVAVGSPLAALVHSTTASGMSGVAGAWLNQAVPDGMFRTRH
                     GLLYTHARGHYHLAVDRPAAALDDFLTCGGLAKEWGMDVPTFLPWRTSAAQAQLALGN
                     PGQARALVREQLGRPGGSTARARGVALRLLAATGEPGRRPALLRESVNLLESCGDQVE
                     LLRALADRCQALDHVGAPAKARMTARHARNVADACGGDALFRKLFRDEVPDREGDPDG
                     LTGEEQGSASLTAAERRVTALAALGHSNREIGRKLFITKSTVEQHLTRVYRKLGVRNR
                     ADLGDLLAGTNLGAGTPAMAMASSAVAG"
CDS                  10933. .11271
                     /codon_start=1
                     /transl_table=11
                     /product="P450-like protein"
                     /protein_id="BAB69191.1"
                     /db_xref="GI:15823974"
                     /translation="MEVGGQTIRAGEGVLCTLSTANRDEGAFTDPDELDLRRDARSHL
                     SFGFGIHQCIGQSLARAEPQIALETLLRRLPGLRLAASFEELRFRDGVGFYGVKELPD
                     PCPALPPDRP"
gene                 11458. .29898
                     /gene="olmA1"
CDS                  11458. .29898
                     /gene="olmA1"
                     /codon_start=1
                     /transl_table=11
                     /product="modular polyketide synthase"
                     /protein_id="BAB69192.1"
                     /db_xref="GI:15823975"
                     /translation="MACRLPGAAGPAEFLALLRNGEDAVGAAPGDRRYAPRRGGFLDS
                     VDEFDAGFFGISPREAAAMDPQQRLVLELSWEALEDAGIPPARLGGSGAGVFVGAIAD
                     DYATLSRAAGADAATPQTTTGLNRGIIANRVSYTLGLRGPSFTVDSGQSSSLVAVHLA
                     AESLRRGECSVALAGGVNLILAEDSTVAVERFGALSPDGYCYTFDARANGYVRGEGGG
                     VVVLKPLADAIADGDDIHCILAGSAVNNDGGGEGLTVPDRAGQEAVLTAAYEQAGISP
                     AAVGYVELHGTGTPTGDPVEAAAVGAVLGAGRPAGQPLLVGSVKTNIGHLEGAAGIAG
                     LLKAVLAVRHREIPATLHHASPNPAIPMAELGIGVNTALTPWPEQAGPLVVGVSSFGM
                     GGTNCHVVLTEPAGYDAAARPATAGIRPAGTPAPLLVSGRDEQALRAQARKLRDHLDT
                     HPGLPLRDVAHTLAAARAPLARRAVLLAADREDVSAGLARLADGTPGADLVRGTVSPG
                     ALAFVFTGQGSQRPGMTAELYRTSAEYAAALDEVCASLDPWLRRPLRDVLFAAEGSDE
                     AALLDRTEFTQPALFAVEVALFRYAEHCGLTPRFLLGHSVGELAAAHVAGVLSLADAC
                     TLVAARGRLMQDRPATGAMVSVQATEDELAPFVSESVTLAALNGPASTVIAGDEDAVL
                     AVAAHWAARGRKTKRLRVSHAFHSPHMDGMLTEFRRIAAGLTFGAPRIPIVSNETGTL
                     LTEADARSPEYWVRHARAAVRPLDGVRLLEDQGVTTLLELGPDGVLSAMARDCLRDAG
                     TVCVPLLRGRDRTEPEALLAALAGLYVRGVPMDWERTATEAGARRVPLPTYAFQRRRH
                     WLPEPSASASRPVVNGGGHAGQAGHAGQAGRAGQAGHAGHDGEPAPAVRDRTTPELVR
                     DRTTLELVRDATAQVLGHASPDDVATHEAFRELGFSSLMLAELGERLTEATGRRVPTT
                     LLFDHPTPDALARRLEADGAERTDAGPVSPPAAAAHDDDPIVIVGMGCRLPGGVRSPE
                     DLWRFVAAGGDAISALPVDRGWEFAADFPGEGGFLADVAEFDAGFFGISPREALAMDP
                     QQRLLLETSWEALERAGVDALKLRGSRTGVFVGASPSEYGPRLHETSDADGHVLTGTA
                     PSVLSGRIAYVLGLEGPALTVDTACSSSLVALHLAAQALRGGECDLALAGGVTVMATS
                     GMFAEFARQGGLAPDGRCKAFADGADGTGWGEGVGMLAVQRLSDAVRDGRPVLAVVRG
                     SAVNSDGASNGLTAPNGPSQQRVIRQALASAGLSAADVDMVEAHGTGTSLGDPIEAQA
                     LLATYGQDRPDDRPLWLGSVKSNIGHTQAAAGVAGVIKSVLALRNGLLPKTLHVDEPT
                     PEVDWSAGAVELLTEGREWPETDGPRRAGVSAFGISGTNAHVILEQAPSVEEPVTAPA
                     DDGLLVPWVVSARSEEALRAQARRLADHVRESGLRPADVGLSLAVTRAGLEHRAVLVA
                     EDEAGFLAHLDALAEGRTPAGVVVEGGTAFLFSGQGSQRAGMGRELYSAHPVFAAAFD
                     EVCAAFDGLLPGALREVVFGGAGLDRTEWAQPALFALEVALFELVSSWGVRADVLLGH
                     SIGELAAAYVAGVWSLADACRIVAARGRLMQALPEGGAMVAVEAAEDELPDLPGGVSV
                     AAVNGPRSLVLSGDEEPVTALAQAFAEQGRRTKQLTVSHAFHSARMEPMLAEFAETLA
                     AVEFRPPRIPVVSNVTGEMAGDEFTTPAYWVRHVREAVRFADGVGAAHAAGVRRYLEL
                     GPDGTLTTLAHDTAHDAGRDVAAVPALRGDQPETRTVVTAVGRLYATGLTVDWPAFFA
                     PYGARPTELPTYAFQHRRYWLEPVAPSSAPPANPLRYRTAWASVPDGPRPSLSGTWAV
                     VVPASGSYEDIVAGVTAALRDHGARPVVVEVDPAGPERADVADTLRAALGEDAGGVVS
                     LLGTAGSLSSGVAATAVLVQAWTDLGGGARLWCLTRGAVSVSPSDPLAAPAQAQLWGM
                     GRVAALEHPERWGGLVDLPAVPDDRAWTRLCAVLAGGSGEDQVAVRAAGLFARRLGRA
                     GTRADAAPAGPSPWRTDGTVLVTGGTGALGAHLAQWLAAAGAAHVLLTSRRGPDADGA
                     AELTARLRETGTEVTVAACDVTDRDALAELLSGLPADRPLTGVVHAAGVLDDGVLDSL
                     TPDRFAAVAGPKVTGAWHLHELTRDLDLSAFVLFSSFAGTVGLAGQANYAAANAYLDA
                     LAVHRAQLGLPATAVAWGSWSGAGMAADTEAARRQLARTGLVPLAPDAALAALGQVMD
                     DGETAVTVADVDWDRFAAGFGPGRPSPLLTGVPEVRRIEERETGRAGAPAETDRLGGL

  Query Match           31.7%;  Score 139;  DB 1;  Length 104326;
  Best Local Similarity   62.0%;  Pred. No. 8.5e-17;
  Matches  220;  Conservative    0;  Mismatches  135;  Indels    0;  Gaps    0;

Qy         21 GCCCGGTGCCGGCAACGTGTCAGCAGCCGGACTCGACCCCACCGAACACCCCCTACTCGG 80
              |||||| |||||| || || | || || || || | |||| | |||||||| || |||||
Db      98184 GCCCGGGGCCGGCGACATGGCCGCGGCGGGGCTGGCCCCCGCGGAACACCCGCTGCTCGG 98125

Qy         81 CGCCACATTGGAACTGGCGACTGACGGTGGAGCGCTTCTTGCAGGGCGCTTGTCTTTGAG 140
              |||    |||  | |||| | |  || ||   |||  |  | ||  |  ||||   |
Db      98124 TGCCGTGCTGGTGCCGGCGGCCGGGGGCGGCCGGCTGTTCACCGGCAGGCTGTCCACGGA 98065

Qy        141 GTCGCATCCGTGGCTGGCTGACCATGCCGTCGGCGGCACGGTGCTGCTGTCGGGCGCCAC 200
                 |  |||||||||| ||||| ||  |||||| ||| || |||||| ||||| || |
Db      98064 CAGCCTGCCGTGGCTGGCGGACCACGCGATCGGCGACACCGTTCTGCTGCCGGGCACCGC 98005

Qy        201 CTTCCTCGAACTCGCCCCTTCATGCGGGCACATACGTGGGCTGCGACCGAGTGGATGAGCT 260
              | ||  |||||| || ||    || || || ||| | |||  ||||   || |  || ||
Db      98004 CGTCGCCGAACTGGCGCTGTGGGCCGGGACGTACCTCGGCCTCGACTCGGTCGGGGATCT 97945

Qy        261 GACGCTGCATGCGCCGCTGGTGGTTCCTGTGGATGGGGGTGTGAGTGTGCAGGTTGGGGT 320
              ||||||  |    ||||| ||  | ||   |   || ||  || | || ||| | |  ||
Db      97944 CACGCTGGAGCTTCCGCTCGTCCTGCCGTCGAGCGGCGGCCTGCGCGTCCAGCTCGCCGT 97885

Qy        321 TGCGGCTGCGGATGGGGAGGGGCGGCGTTTGGTGAGTGTGTATGCGCGGGGTGGG 375
              |  ||| |||| | |   ||   ||||  | |      |||| || | |||  | |
Db      97884 GGGTGCTCCGGACGAGTCCGGCGAGCGTGAGTTCGCCCTGTACGCACAGGCCGAG 97830


RESULT 11
SNA278573/c
LOCUS       SNA278573              84985 bp    DNA     linear   BCT 17-MAY-2001
DEFINITION  Streptomyces natalensis pimaricin biosynthetic gene cluster.
ACCESSION   AJ278573
VERSION     AJ278573.1  GI:12055067
KEYWORDS    ABC transporter; cholesterol oxidase; cytochrome P450 monoxygenase;
            ferredoxin; glucosyl transferase; macrolide efflux pump; mycosamine
            dehyralase; orfX; pimA gene; pimB gene; pimC gene; pimD gene; pimE
            gene; pimF gene; pimG gene; pimH gene; pimI gene; pimJ gene; pimK
            gene; pimS0 gene; pimS1 gene; pimS2 gene; pimS3 gene; pimS4 gene;
            polyketide synthase; sensory transduction protein; thioesterase.
SOURCE      Streptomyces natalensis.
  ORGANISM  Streptomyces natalensis
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1  (bases 1 to 84985)
  AUTHORS   Aparicio,J.F., Fouces,R., Mendes,M.V., Olivera,N. and Martin,J.F.
  TITLE     A complex multienzyme system encoded by five polyketide synthase
            genes is involved in the biosynthesis of the 26-membered polyene
            macrolide pimaricin in Streptomyces natalensis
  JOURNAL   Chem. Biol. 7 (11), 895-905 (2000)
  MEDLINE   20547809
   PUBMED   11094342  -
REFERENCE   2  (bases 1 to 84985)
  AUTHORS   Aparicio,J.F.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-JUN-2000) Aparicio J.F., Molecular Genetics,
            Institute Of Biotechnology - Inbiotec, AV. REAL 1 - LEON, 24006,
            SPAIN
FEATURES             Location/Qualifiers
     source          1. .84985
```

```
                /organism="Streptomyces natalensis"
                /db_xref="taxon:68242"
CDS             <1. .2086
                /note="ORFX"
                /codon_start=2
                /transl_table=11
                /product="hypothetical protein"
                /protein_id="CAC20917.1"
                /db_xref="GI:12055068"
                /translation="TRPDAFREATEPYGALLAPLMPERSADRVEGTKWTSDWSQARFL
                AHDAVCEVLLTLAGQHPLVLLLEDLQWADTASLDLLRLLSPRCQGHPLGIVLTARKFE
                TESDATLRRIVSDVLRGPRTATLRLCGLPRQAVGALVEAHAGSGVDPEVVEVLHRRSE
                GNPYFVMQLLSLVGDARRLHRPDAVDALLAHIPTGVREALNQRFAALPEPVLRVLRLC
                AVIGAEVDTDLLRRTATCDASVNAGLESAIRAGLLGEDPHHPGQLHFAHALVQETLFD
                ELPRADRRRLHAKVADALYARRRGHVGDEEIERVAHHSWHAKSALPAAEVLPRLLLAA
                EHAEQAMAYEQVEIWLRRAVHLVGFLPPDDPATVRLEQKLHIQLGQVLATTRGYGHPE
                AETALTRGRALSTVTHAPEDPSVLWALCAASLVTGRYDASRQLSGLLRDLSWQTREPV
                AVLGAAYGEGIVLHVRGMLPQALAELEHGVERADHFANEGHALARTFQHDPRVSCRSY
                NTFTHWLLGHRQIADARRRELLSLTEYESRPSDRSFALYVDGVVAAWEGDTRTALASS
                DEGVRVAGDHGLLYWKAMLGVVKGWGLVHAGDRDEGLALMHTSLAELGQSRTYLRHPL
                HLGLLGQAQHHNGQIEEARTTFRMLLAAVEQRRERVYLHPALPATPLLHELLGHGADE
                ALADRDPGRGQV"
gene            2190. .3566
                /gene="pimK"
CDS             2190. .3566
                /gene="pimK"
                /function="putative glucosyl transferase"
                /codon_start=1
                /transl_table=11
                /product="PimK protein"
                /protein_id="CAC20918.1"
                /db_xref="GI:12055069"
                /translation="MESARRPILFVSLPESGLLNPLLVLAGELSRQGVEDLWFATDEP
                RRNDVKRIAEGSPVEFASLGEVDSEMSAVTWSDEVYREVTQPSRFKAHRAVVRHTYRP
                GLQAEKFRRLQAVIDEVQPALMVIDCISGFAVDAAIARNIPYVLSVPFLPSNVLTAHT
                HFAKSYTPRGFPVPHTGLSRRMTLAQRVANELFKLRTFAMFLNPRLGKVLAEDNRRRN
                ELGLPKASFMARIEHADLVLCNSLAELDYPFDIPEKMRLVGAMVPPLPEAPDDQDLSR
                WLDAQSSVVYVGLGTITRLTREQVGSMVEVARRLEDRHQVLWKLPSEQQHLLPPRESL
                PGNLRVESWVPSQMDVLAHPHVKVFFTHGGGNGFNEGMYFGKPLVVRPLWVDCYDQAV
                RGQDFGLSLTLDRPQTIDVNDVVDKLTRVLGTPSFYEKAERRAALMRSAGGRETAAGL
                VLSLPALA"
gene            complement(3662. .9736)
                /gene="pimS4"
CDS             complement(3662. .9736)
                /gene="pimS4"
                /function="polyketide synthase"
                /codon_start=1
                /transl_table=11
                /product="PimS4 protein"
                /protein_id="CAC20919.1"
                /db_xref="GI:12055070"
                /translation="MPNDKPNDKKLVDYLKWVTNDLHQTRQRLREVESAKQEPIAIVG
                MSCRLPGGVRSPEEFWQLLKDGRDGISTFPTDRGWDLDLLAGDGSGSSATAEGGFIDA
                ASFDAGFFGISPREAVTMDPQQRILLESSWEALERAGIDPVSLRGTETGVFVGTSGID
                YINTVMNSREDIEGHGSTGLIASILSGRVSYTFGLEGPAVTVDTACSSSLVSLHLAAQ
                ALRVGECSLALAGGVTAISTPMSFVGFSRQGGLAPDGRCKAFSDTADGTAWSEGVGVL
                VVERLSDARRNGHPVLAVVRGSAVNQDGASNGLTAPSGPSQQRVIRQALAAAGLSASD
                VDVVEAHGTGTALGDPIEAQALLATYGRDRDPERPLLLGSVKSNLGHTQAAAGAAGLI
                KMILAMRHGVLPRTLHVEAPSSHVDWSSGALQLLTEPQEWPKADRPWRAGVSSFAISG
                TNAHIILEQPPAAEEPEDEEGPHTSPSAVPWLVSAKSEEALTAQLDRISALTDRSPLD
                LGLSLATTRAHLEHRAVLLSEGGETVEVARGHAEHTAGKLAVLFSGQGAQRVGMGREL
                YGRFPVFAEALDGVLAHFDAGLRDVMFGEAEGLDETGFTQPALFAIEVALFRLSESLG
                VRPDFVAGHSIGEIAAAHVAGVFSLADACALVAARARLMQALPAGGAMVAVQAAESEV
                AARLVDGVSIAAVNGPEAVVIAGEEREVLRIAEEFTGLGRKTRRLPVSHAFHSPLMEP
                MLEDFRRVAEGLSYEAPRIALVSNLTGELATEDLVRSAEYWVRHVRETVRFADGVSAL
                AAEGASVFLELGPDGVLTAMAQHTLDGTATTVPALRKDRSEETALLTALAQLHVAGVG
                VDWSGIFAGTGARRVDLPTYPFQHERYWPRPAVLSGDVVGAGLRSAEHPLLGATLTLA
                ESGEVVLTGRLSLQAYPWLTDHTMGGSVLFPATGFLELAMRAADQVGCDRVDEFVLMT
                PLVLAEESGTQVQVLVGPADEAGTRTISIHARPDGDPDQPWMQHAGGTLTSGERIVDF
                GSTVWPPQGAEAVDLEGFYDTTGYGPAFQGLRSVWQCEDGAYVEVALPDQVAEDAGAF
                GLHPVLLDAVLQSRRMAGIGAAGDTFLPFAWRGVSLHAGGASVLRARVIKTGEDTLSL
                VAVDAEGAPVLSVEELVLRTQLTHTDAPLTARRGAERDALLSLEWVSGPEVPPAGVRC
                VSLGADELGVGASVATLADLSGDEDFVVVPVSGVGGDVPAAAHELTARTLGLLQEWSG
                REGTSASRLVFVTRNAVSADGGESVRDVAAAAVWGLVRSAQSESPGRIALVDLEADSG
                DLAEVLAGLPGLLAAGDAQFVVREGAVRVGRLVESNEGAGVSPREWDADGTVLVTGGT
                GGLGRELARHLVVERGVRHLLLVSRSGPEASGVGELRDELAAHGAEIAVRACDVADRA
                AVEAVVGAVDAAHPLTAVIHTAGVLDDGVIASLTPERLSAVLRPKVDAAWNLHEATKD
                LDLAAFVLYSSISGVLGSLGQGNYAAGNVFLDALAGHRAGLGLPAQSLAWGAWTQGGG
                MTATLSDADMQRMASYGTAPLSPERGLALFDAATAGSATHLVAVGRVSGPVRAQGPVP
                SMLRGLVKGARRTAVNTTGETAVSLADRLRAARPEERTQLMTDVVRTEAAAVLGHASS
                QNVDARREFYELGFDSLTSVELRNRLSTVTGLRLSATVVFDSKNPTDLAARLYADLAA
                QSALDGSASQAGFLPVAAPETDSLERMFLDALDSGKVPEAQRMLSALGALRPSFENTA
                ELEDLPLPATLAEGPGTPRLICVSTPTANGGVHEYARLAASFRGERHVSALPLVGPAT
                GERLPATAETAVRVIAESALRASDGNPFVLVGHSSAGAFAYLAAALLENTWGIRPEAV
                VLLDTLSLRHEQNESIDYAGLMRRHFMVDEVSPVRMTNSRLSAMARWMGMLNQLEVRH
                TTVPVLIIRATKETFGIGTDTGIYGEDHGSPVDVRSVDADHFSMVRDDAPETARIVRE
                WLDSLGNA"
gene            complement(9772. .15198)
                /gene="pimS3"
CDS             complement(9772. .15198)
                /gene="pimS3"
                /function="polyketide synthase"
                /codon_start=1
                /transl_table=11
                /product="PimS3 protein"
                /protein_id="CAC20920.1"
                /db_xref="GI:12055071"
                /translation="MTTSSETVVAALRASLKEAENLRRQNRKLVAAATEPIAIVAMAC
                RYPGGVQSPEDLWELLTAGKDAIGAFPTDRGWDLDALRDAGVDRRGNAVSQEGGFLDG
                VGDFDADFFGISPRESVTMDPQQRLLLETSWEAIERAGIDPTTLRGSRTGVFIGTNGQ
                DYAYLLVRSLADATGDIGTGIAASATSGRLSYTLGLEGPAVTVDTACSSSLVALHTAS
                HALRAGECSLALVGGVNVMSSPGSLMEFSRQGGLAADGRCKAFSDTADGTGWSEGVGV
                LVVERLSDARRNGHPVLAVVRGSAVNQDGASNGFTAPSGPSQQRVIRQALAAAGLSAS
                DVDVVEAHGTGTPLGDPIEAQALLATYGQDRDPERPLLLGSVKSNLGHTQAAAGAAGL
                IKMILAMRHGVLPKTLYADVPSSHVDWTSGEVELLTEPQEWPKADRPWRAGVSSFGIS
                GTNAHVIVEQAEPDEEPVAEPGLAPAVVPWLVSAKTEEALAAQLARISSRRGTASPVD
                VGFSLASGRSAFEHRAVLLAQCPDGELSEAARGRAETSRPLAVLFSGQGAQRVGMGRE
                LYVRFPVFAEALDEVLAHFDAGLRDVMFGDAEGLDETGFTQPALFAIEVALFRLAESL
                GVRPDFVGGHSIGEIAAAHVAGVFSLTDACALVAARARLMQALPAGGAMVAVQASESE
                VAARLVDGVSIAAVNGPEAVVLAGDETAVLRIAEELAAGGRKTQRLPVSHAFHSPLME
                PMLEDFRRVAEGLSYGAPQIPLVSNVTGELAVEELVCSPDYWVRHVRETVRFADGVSA
                LEAEGASAFLELGPDGVLTAMARHTLDETKVAVPVLRKDRPEETALLTALAHLHVAGI
                GVDWSGIFAGTGARRVDLPTYPFQHERYWPRPAVLSGDVAGAGLRSAEHPLLGATLTL
                AESGEVVLTGRLSLQTHPWLTDHAIGGNVFFPGTGFLELAMHAADQAGCDRVEDLVLM
                TPLVLAEENGTQVQVLVGPADEAGTRTISIHARPEDDPDQPWMQHAGGTLTSGERVVD
                FGSTVWPPQDAVAVDLEGFYEGTGYGPAFQGLRSVWSCGDGAYAEVALPDHVAEDAGA
                FGLHPALLDAVLQSHRIAGVGSADDQFLPFAWRGVSLHAGGASVLRVRVVKTGDDSMS
                IAAVDAEGAPVLSAEAMVMRAYRDPGTGQAAARHGAERDELLSLEWVSGPEVPPAGVR
                CVSLGADELGVGASVASLADLSGDEDFVVLPVSGVGGDVPAAAHELTARTLGLLQEWS
                GREGSSGSRLVFVTRNAVSADEGESVRDVAAAAVWGLVRSAQSESPGRIALVDLEADS
                GDLAEVLAGLPGLLAAGDAQFVVREGAVRVGRLVESNEGAGVSPREWDADGTVLVTGG
                TGGLGRELARHLVVERGVRHLLLVSRSGPEAPGVSGLSDELAAHGAEVTVRACDVADR
                AAVHALVAAVPAEHPLTAVIHTAGVLDDGLVTSLTPERLSTVLRPKVDAAWNLHEATK
                DLDLAAPVLYSSISGVLGAPGQANYAAGNVFLDALAWHRMGQGLPAQSLAWGAWTQDG
                GMTATLSDTDMQRMASYGTAPLSRQRGLALFDAACARSAAHLVAVGRVGGPVRASAPV
                PSVLRGLVKGARRTAVNTTGETAVSLADRLRAARPEERPRLMTDLVRTEAAAVLGHAS
                AKNVDARREFNELGFDSLTSVELRNRLSALTGLRLSAALVFDYPNPTALADHFVSELV
                EEAATGPALLAELERLDSALAAGAPDARTRAAVTARLAQILDDWRDAPAEEDGAEVAE
                RLEAASTDEIFAFIDNELGRHSDR"
gene            complement(15210. .43733)
                /gene="pimS2"
CDS             complement(15210. .43733)
                /gene="pimS2"
                /function="polyketide synthase"
                /codon_start=1
                /transl_table=11
                /product="PimS2 protein"
                /protein_id="CAC20921.1"
                /db_xref="GI:12055072"
                /translation="MDQEQKLRDYLKRASADLRRSRQRVGELEAAAREPIAIVGMSCR
                FPGGVRSPEELWSLLAADGDGITKSPADRGWELLAEDAGEFSGGFLHDAAEFDASFFG
                ISPREALAMDPQQRLLLESSWEAFERAGIDPAAVRGTRTGMFVGAMPQEYRVGPDDDV
                QGFALTGTTTSVISGRLAYFFGAVGPAVTVDTACSSSLVALHLAAHSLRQGECSLALA
                AGVTVMSSPTTFVEFNRQGGLSADGRCRSFADSADGTGWSEGVGVLVLERLSEARRNG
                HEILAVIRGSAVNQDGASNGLTAPNGPSQRRVIEQALVSARLSADEVDVVEAHGTGTT
                LGDPVEAQALLATYGQGRDDDRPLLLGSVKSNLGHTQAAAGMAGVIKMVLAMRHGVLP
```

```
                     RTLHVDAPSSHVDWTQGAVRLLTEHVPWPQGEQPRRAGVSSFGLSGTNAHTILEEAPA
                     APSPEAGQEDTAPVAEPLAAGAVPWLVSGRTREALRAQAARLLEHLTSRPGLAPLDVA
                     YSLATARSGLEHRAAFTAAGPDGARAALTALADGTPAPGLVQDTARTRSKLAFLFAGQ
                     GSQRPGMGRELAARFPVFATALDEVLGHFDEGLERPLKDVLFAAEGTPEAALLDQTGY
                     AQPALFAIEVALYRLAESFGITPDFLAGHSIGEIAAAHVAGVFTLADAAALVLARGRL
                     MQALPEGGAMVSLEATEEEVLPLLEDRPDRLSIAAVNGPAAVVIAGEADDVTAVAEHF
                     DALGRRTKRLRVSHAFHSPLMEPMLAEFRTAVSRLTPQAPVIPVVSGLTGTLATVEQL
                     TSPDYWLDHARHAVRFADALRLARRARRRHLPRMGPDGVLSAMAQSCCRGHGHRRRRR
                     LRHARRVRPGRPEAETLTAALAGLHTRGVAVQWEPYFQGTGARRADLPTYAFQRRRYW
                     PKSLPATGGDVRAAGLGAAHHPLLTAAVSVANSDGLLLTGRLSRRTHPWLADHAVRGT
                     VLLPGTAFLELAVRAGDEAGCGRVEELTLAAPLVLPEEGGVQVQVWVGSPDASGRRAV
                     SVHSRPDGPEELPWTQHAAGTLVAGEHHAGFDATVWPPADATPLDLDGFYARMADTGF
                     GYGPLFQGLRAAWRSGDDVYAEVALPDSDGAASTESSAAFGLHPALLDAALHAAAFVD
                     LGEDARGGLPFSWQDVTLHASGASTVRVKLTPDGDDAVALAVADTTGAPVASIGSLVL
                     RTLPDEQIDAAHSLVRDALFELRWSEARQTAHEAPASAAVLGPDLFGLERALAPTDLT
                     VTAVPGPDAAPHAPDVVLAPVAGTTEDDVVTSVHARTAHVLERLRTWLTDERYAASRL
                     VFVTRGALATDARQASDPVAAAVWGLVRAAQSEHPGRFGLLDLDPGTTEPATEPLLRA
                     LALSSDEPQTAVRGTAVLTARLARAQAPQAATEWDPEGTVVLTGATGGLGRVLARHLV
                     AERGVRHLLLVSRRGADADGAGELMAELAAHGAQVTLAACDLADRAAVRRLLSGVPAE
                     HPVTAVVHSAGSLDDGVIASLTPERISAVLRPKADAVWHLHEATRDLDLAAFVVFSSL
                     SGTVGAAGQGNYAAANAFLDALAQLRRAQGLPGLSLAWGPWAPTAGMTGGLTDDDLDR
                     LARMGTPALTEEQGIALFDAATATDAAVLLPTRIDLSVLRVQSEIPPLWRGLIRTPAR
                     RSVVSASETAVTLVQQLSRLAETDRREVVLDLVAGQVAAVLGHAGTADIDPRRPLREL
                     GFDSLTAVELRNRLSAATGLRTAATVIFDHPTVDALAAHLLDGLMGSEAAAAADAPRT

  Query Match             31.1%;  Score 136;  DB 1;  Length 84985;
  Best Local Similarity   60.6%;  Pred. No. 3.2e-16;
  Matches  223;  Conservative    0;  Mismatches  145;  Indels    0;  Gaps    0;

Qy          1 TACTGGCTCGAAAGCACACAGCCCGGTGCCGGCAACGTGTCAGCAGCCGGACTCGACCCC 60
              ||||||| | | |||   |   |    | |||  |||||   || ||||| || | | |
Db      40997 TACTGGCCCAAGAGCCTGCCCGCGACGGGCGGTGACGTGCGTGCCGCCGGTCTGGGCGCG 40938

Qy         61 ACCGAACACCCCCTACTCGGCGCCACATTGGAACTGGCGACTGACGGTGGAGCGCTTCTT 120
                | | |||||  | ||   ||  |  ||    | || |    ||  ||    || ||
Db      40937 GCACACCACCCGTTGCTGACGGCGGCCGTGTCGGTCGCCAACTCCGACGGGCTCCTGCTG 40878

Qy        121 GCAGGGCGCTTGTCTTTGAGGTCGCATCCGTGGCTGGCTGACCATGCCGTCGGCGGCACG 180
              | || ||  ||||   | |  | || || ||||| || ||||| |||||  ||||||||
Db      40877 ACCGGACGGCTGTCGCGGCGTACCCACCCCTGGCTCGCCGACCACGCCGTGCGCGGCACG 40818

Qy        181 GTGCTGCTGTCGGGCGCCACCTTCCTCGAACTCGCCCTTCATGCGGGCACATACGTGGGC 240
              ||||||||| ||||| || ||||||| || || ||| | |  || ||      | |  |||
Db      40817 GTGCTGCTGCCGGGCACCGCCTTCCTGGAGCTGGCCGTACGCGCCGGGGACGAGGCCGGC 40758

Qy        241 TGCGACCGAGTGGATGAGCTGACGCTGCATGCGCCGCTGGTGGTTCCTGTGGATGGGGGT 300
              ||||  ||| ||||| || || || ||    || ||||||||| | || | ||| || ||
Db      40757 TGCGGCCGCGTGGAGGAACTCACCCTCGCCGCCCCGCTGGTGCTGCCCGAGGAGGGCGGC 40698

Qy        301 GTGAGTGTGCAGGTTGGGGTTGCGGCTGCGGATGGGGAGGGGCGGCGTTTGGTGAGTGTG 360
              ||    ||||||||  |||| |     | || | |  ||| || ||    || || ||
Db      40697 GTCCAGGTGCAGGTGTGGGTGGGCAGCCCCGACGCGTCGGGCCGCCGCGCCGTCAGCGTT 40638

Qy        361 TATGCGCG 368
                | | ||
Db      40637 CACTCCCG 40630


RESULT 12
AF079138
LOCUS       AF079138               37948 bp    DNA     linear   BCT 30-OCT-1998
DEFINITION  Streptomyces venezuelae methymycin/pikromycin polyketide synthase
            gene cluster, complete sequence.
ACCESSION   AF079138
VERSION     AF079138.1  GI:3808326
KEYWORDS    .
SOURCE      Streptomyces venezuelae.
  ORGANISM  Streptomyces venezuelae
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1  (bases 1 to 37948)
  AUTHORS   Xue,Y., Zhao,L., Liu,H.W. and Sherman,D.H.
  TITLE     A gene cluster for macrolide antibiotic biosynthesis in
            Streptomyces venezuelae: architecture of metabolic diversity
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 95 (21), 12111-12116 (1998)
  MEDLINE   98445333
   PUBMED   9770448
REFERENCE   2  (bases 1 to 37948)
  AUTHORS   Xue,Y., Wilson,D., Zhuo,L. and Sherman,D.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-JUL-1998) Department of Microbiology, University of
            Minnesota, 420 Delaware Street SE 1060, P.O. Box 196, Minneapolis,
            MN 55455, USA
REFERENCE   3  (bases 1 to 37948)
  AUTHORS   Xue,Y., Wilson,D., Zhuo,L. and Sherman,D.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-OCT-1998) Department of Microbiology, University of
            Minnesota, 420 Delaware Street SE 1060, P.O. Box 196, Minneapolis,
            MN 55455, USA
  REMARK    Sequence update by submitter
COMMENT     On Oct 30, 1998 this sequence version replaced gi:3800831.
FEATURES             Location/Qualifiers
     source          1. .37948
                     /organism="Streptomyces venezuelae"
                     /strain="ATCC15439"
                     /db_xref="ATCC:15439"
                     /db_xref="taxon:54571"
     gene            154. .1122
                     /gene="pikR2"
     CDS             154. .1122
                     /gene="pikR2"
                     /note="mls type resistance protein"
                     /codon_start=1
                     /transl_table=11
                     /product="rRNA methyltransferase PikR2"
                     /protein_id="AAC69327.1"
                     /db_xref="GI:3800832"
                     /translation="MAFSPQGGRHELGQNFLVDRSVIDEIDGLVARTKGPILEIGPGD
                     GALTLPLSRHGRPITAVELDGRRAQRLGARTPGHVTVVHHDFLQYPLPRNPHVVVGNV
                     PFHLTTAIMRRLLDAQHWHTAVLLVQWEVARRRAGVGGSTLLTAGWAPWYEFDLHSRV
                     PARAFRPMPGVDGGVLAIRRRSAPLVGQVKTYQDFVRQVPTGKGNGLKEILRRTGRIS
                     QRDLATWLRRNEISPHALPKDLKPGQWASLWELTGGTADGSFDGTAGGGAAGSHGAAR
                     VGAGHPGGRVSASRRGVPQARRGRGHAVRSSTGTEPRWGRGRAESA"
     gene            complement(1283. .2293)
                     /gene="pikR1"
     CDS             complement(1283. .2293)
                     /gene="pikR1"
                     /note="mls type resistance protein"
                     /codon_start=1
                     /transl_table=11
                     /product="rRNA methyltransferase PikR1"
                     /protein_id="AAC69328.1"
                     /db_xref="GI:3800833"
                     /translation="MAMRDSIPRRADRDTLRRELGQNFLQDDRAVRNLVTHVEGDGRN
                     VLEIGPGKGAITEELVRSFDTVTVVEMDPHWAAHVRRKFEGERVTVFQGDFLDFRIPR
                     DIDTVVGNVPFGITTQILRSLLESTNWQSAALIVQWEVARKRAGRSGGSLLTTSWAPW
                     YEFAVHDRVRASSFRPMPRVDGGVLTIRRRPQPLLPESASRAFQNFAEAVFTGPGRGL
                     AEILRRHIPKRTYRSLADRHGIPDGGLPKDLTLTQWIALFQASQPSYAPGAPGTRMPG
                     QGGGAGGRDYDSETSRAAVPGSRRYGPTRGGEPCAPRAQVRQTKGRQGARGSSYGRRT
                     GR"
     gene            2912. .16753
                     /gene="pikAI"
     CDS             2912. .16753
                     /gene="pikAI"
                     /note="multifunctional polyketide synthase; contains a
                     loading module, Pik module 1 and module 2"
                     /codon_start=1
                     /transl_table=11
                     /product="type I polyketide synthase PikAI"
                     /protein_id="AAC69329.1"
                     /db_xref="GI:3800834"
                     /translation="MSSAGITRTGARTPVTGRGAAAWDTGEVRVRRGLPPAGPDHAEH
                     SFSRAPTGDVRAELIRGEMSTVSKSESEEFVSVSNDAGSAHGTAEPVAVVGISCRVPG
                     ARDPREFWELLAAGGQAVTDVPADRWNAGDFYDPDRSAPGRSNSRWGGFIEDVDRFDA
                     AFFGISPREAAEMDPQQRLALELGWEALERAGIDPSSLTGTRTGVFAGAIWDDYATLK
                     HRQGGAAITPHTVTGLHRGIIANRLSYTLGLRGPSMVVDSGQSSSLVAVHLACESLRR
```

```
GESELALAGGVSLNLVPDSIIGASKFGGLSPDGRAYTFDARANGYVRGEGGGFVVLKR
LSRAVADGDPVLAVIRGSAVNNGGAAQGMTTPDAQAQEAVLREAHERAGTAPADVRYV
ELHGTGTPVGDPIEAAALGAALGTGRPAGQPLLVGSVKTNIGHLEGAAGIAGLIKAVL
AVRGRALPASLNYETPNPAIPFEELNLRVNTEYLPWEPEHDGQRMVVGVSSFGMGGTN
AHVVLEEAPGGCRGASVVESTVGGSAVGGGVVPWVVSAKSAAALDAQIERLAAFASRD
RTDGVDAGAVDAGAVDAGAVARVLAGGRAQFEHRAVVVGSGPDDLAAALAAPEGLVRG
VASGVGRVAFVFPGQGTQWAGMGAELLDSSAVFAAAMAECEAALSPYVDWSLEAVVRQ
APGAPTLERVDVVQPVTFAVMVSLARVWQHHGVTPQAVVGHSQGEIAAAYVAGALSLD
DAARVVTLRSKSIAAHLAGKGGMLSLALSEDAVLERLAGFDGLSVAAVNGPTATVVSG
DPVQIEELARACEADGVRARVIPVDYASHSRQVEIIESELAEVLAGLSPQAPRVPFFS
TLEGAWITEPVLDGGYWYRNLRHRVGFAPAVETLATDEGFTHFVEVSAHPVLTMALPG
TVTGLATLRRDNGGQDRLVASLAEAWANGLAVDWSPLLPSATGHHSDLPTYAFQTERH
WLGEIEALAPAGEPAVQPAVLRTEAAEPAELDRDEQLRVILDKVRAQTAQVLGYATGG
QIEVDRTFREAGCTSLTGVDLRNRINAAFGVRMAPSMIFDFPTPEALAEQLLLVVHGE
AAANPAGAEPAPVAAAGAVDEPVAIVGMACRLPGGVASPEDLWRLVAGGGDAISEFPQ
DRGWDVEGLYHPDPEHPGTSYVRQGGFIENVAGFDAAFFGISPREALAMDPQQRLLLE
TSWEAVEDAGIDPTSLRGRQVGVFTGAMTHEYGPSLRDGGEGLDGYLLTGNTASVMSG
RVSYTLGLEGPALTVDTACSSSLVALHLAVQALRKGEVDMALAGGVAVMPTPGMFVEF
SRQRGLAGDGRSKAFAASADGTSWSEGVGVLLVERLSDARRNGHQVLAVVRGSALNQD
GASNGLTAPNGPSQQRVIRRALADARLTTSDVDVVEAHGTGTRLGDPIEAQALIATYG
QGRDDEQPLRLGSLKSNIGHTQAAAGVSGVIKMVQAMRHGLLPKTLHVDEPSDQIDWS
AGAVELLTEAVDWPEKQDGGLRRAAVSSFGISGTNAHVVLEEAPVVVEGASVVEPSVG
GSAVGGGVTPWVVSAKSAAALDAQIERLAAFASRDRTDDADAGAVDAGAVAHVLADGR
AQFEHRAVALGAGADDLVQALADPDGLIRGTASGVGRVAFVFPGQGTQWAGMGAELLD
SSAVFAAAMAECEAALSPYVDWSLEAVVRQAPGAPTLERVDVVQPVTFAVMVSLARVW
QHHGVTPQAVVGHSQGEIAAAYVAGALPLDDAARVVTLRSKSIAAHLAGKGGMLSLAL
NEDAVLERLSDFDGLSVAAVNGPTATVVSGDPVQIEELAQACKADGFRARIIPVDYAS
HSRQVEIIESELAQVLAGLSPQAPRVPFFSTLEGTWITEPVLDGTYWYRNLRHRVGFA
PAIETLAVDEGFTHFVEVSAHPVLTMTLPETVTGLGTLRREQGGQERLVTSLAEAWVN
GLPVAWTSLLPATASRPGLPTYAFQAERYWLENTPAALATGDDWRYRIDWKRLPAAEG
SERTGLSGRWLAVTPEDHSAQAAAVLTALVDAGAKVEVLTAGADDDREALAARLTALT
TGDGFTGVVSLLDGLVPQVAWVQALGDAGIKAPLWSVTQGAVSVGRLDTPADPDRAML
WGLGRVVALEHPERWAGLVDLPAQPDAAALAHLVTALSGATGEDQIAIRTTGLHARRL
ARAPLHGRRPTRDWQPHGTVLITGGTGALGSHAARWMAHHGAEHLLLVSRSGEQAPGA
TQLTAELTASGARVTIAACDVADPHAMRTLLDAIPAETPLTAVVHTAGALDDGIVDTL
TAEQVRRAHRAKAVGASVLDELTRDLDLDAFVLFSSVSSTLGIPGQGNYAPHNAYLDA
LAARRRATGRSAVSVAWGPWDGGGMAAGDGVAERLRNHGVPGMDPELALAALESALGR
DETAITVADIDWDRFYLAYSSGRPQPLVEELPEVRRIIDARDSATSGQGGSSAQGANP
LAERLAAAAPGERTEILLGLVRAQAAAVLRMRSPEDVAADRAFKDIGFDSLAGVELRN
RLTRATGLQLPATLVFDHPTPLALVSLLRSEFLGDEETADARRSAALPATVGAGAGAG
AGTDADDDPIAIVAMSCRYPGDIRSPEDLWRMLSEGGEGITPFPTDRGWDLDGLYDAD
PDALGRAYVREGGFLHDAAEFDAEFFGVSPREALAMDPQQRMLLTTSWEAFERAGIEP
ASLRGSSTGVFIGLSYQDYAARVPNAPRGVEGYLLTGSTPSVASGRIAYTFGLEGPAT
TVDTACSSSLTALHLAVRALRSGECTMALAGGVAMMATPHMFVEFSRQRALAPDGRSK
AFSADADGFGAAEGVGLLLVERLSDARRNGHPVLAVVRGTAVNQDGASNGLTAPNGPS
QQRVIRQALADARLAPGDIDAVETHGTGTSLGDPIEAQGLQATYGKERPAERPLAIGS
VKSNIGHTQAAAGAAGIIKMVLAMRHGTLPKTLHADEPSPHVDWANSGLALVTEPIDW
PAGTGPRRAAVSSFGISGTNAHVVLEQAPDAAGEVLGADEVPEVSETVAMAGTAGTSE
VAEGSEASEAPAAPGSREASLPGHLPWVLSAKDEQSLRGQAAALHAWLSEPAADLSDA
DGPARLRDVGYTLATSRTAFAHRAAVTAADRDGFLDGLATLAQGGTSAHVHLDTARDG
TTAFLFTGQGSQRPGAGRELYDRHPVFARALDEICAHLDGHLELPLLDVMFAAEGSAE
AALLDETRYTQCALFALEVALFRLVESWGMRPAALLGHSVGEIAAAHVAGVFSLADAA
RLVAARGRLMQELPAGGAMLAVQAAEDEIRVWLETEERYAGRLDVAAVNGPEAAVLSG
DADAAREAEAYWSGLGRRTRALRVSHAFHSAHMDGMLDGFRAVLETVEFRRPSLTVVS
NVTGLAAGPDDLCDPEYWVRHVRGTVRFLDGVRVLRDLGVRTCLELGPDGVLTAMAAD
GLADTPADSAAGSPVGSPAGSPADSAAGALRPRPLLVALLRRKRSETETVADALGRAH
AHGTGPDWHAWFAGSGAHRVDLPTYSFRRDRYWLDAPAADTAVDTAGLGLGTADHPLL
GAVVSLPDRDGLLLTGRLSLRTHPWLADHAVLGSVLLPGAAMVELAAHAAESAGLRDV
RELTLLEPLVLPEHGGVELRVTVGAPAGEPGGESAGDGARPVSLHSRLADAPAGTAWS
CHATGLLATDRPELPVAPDRAAMWPPQGAEEVPLDGLYERLDGNGLAFGPLFQGLNAV
WRYEGEVFADIALPATTNATAPATANGGGSAAAAPYGIHPALLDASLHAIAVGGLVDE
PELVRVPFHWSGVTVHAAGAAAARVRLASAGTDAVSLSLTDGEGRPLVSVERLTLRPV
TADQAAASRVGGLMHRVAWRPYALASSGEQDPHATSYGPTAVLGKDELKVAAALESAG
VEVGLYPDLAALSQDVAAGAPAPRTVLAPLPAGPADGGAEGVRGTVARTLELLQAWLA
DEHLAGTRLLLVTRGAVRDPEGSGADDGGEDLSHAAAWGLVRTAQTENPGRFGLLDLA
DDASSYRTLPSVLSDAGLRDEPQLALHDGTIRLARLASVRPETGTAAPALAPEGTVLL
TGGTGGLGGLVARHVVGEWGVRRLLLVSRRGTDAPGADELVHELEALGADVSVAACDV
ADREALTAVLDAIPAEHPLTAVVHTAGVLSDGTLPSMTTEDVEHVLRPKVDAAFLLDE
LTSTPAYDLAAFVMFSSAAAVFGGAGQGAYAAANATLDALAWRRRAAGLPALSLGWGL
WAETSGMTGELGQADLRRMSRAGIGGISDAEGIALLDAALRDDRHPVLLPLRLDAAGL
RDAAGNDPAGIPALFRDVVGARTVRARPSAASASTTAGTAGTPGTADGAAETAAVTLA
DRAATVDGPARQRLLLEFVVGEVAEVLGHARGHRIDAERGFLDLGFDSLTAVELRNRL
NSAGGLALPATLVFDHPSPAALASHLDAELPRGASDQDGAGNRNGNENGTTASRSTAE
```

```
TDALLAQLTRLEGALVLTGLSDAPGSEEVLEHLRSLRSMVTGETGTGTGTASGAPDGAGS
GAEDRPWAAGDGAGGGSEDGAGVPDFMNASAEELFGLLDQDPSTD"
gene            16858. .28077
                /gene="pikAII"
CDS             16858. .28077
                /gene="pikAII"
                /note="multifunctional polyketide synthase; contains Pik
                module 3 and module 4"
                /codon_start=1
                /transl_table=11
                /product="type I polyketide synthase PikAII"
                /protein_id="AAC69330.1"
                /db_xref="GI:3800835"
                /translation="MSTVNEEKYLDYLRRATADLHEARGRLRELEAKAGEPVAIVGMA
                CRLPGGVASPEDLWRLVAGGEDAISEFPQDRGWDVEGLYDPNPEATGKSYAREAGFLY
                EAGEFDADFFGISPREALAMDPQQRLLLEASWEAFEHAGIPAATARGTSVGVFTGVMY
                HDYATRLTDVPEGIEGYLGTGNSGSVASGRVAYTLGLEGPAVTVDTACSSSLVALHLA
                VQALRKGEVDMALAGGVTVMSTPSTFVEFSRQRGLAPDGRSKSFSSTADGTSWSEGVG
                VLLVERLSDARRKGHRILAVVRGTAVNQDGASSGLTAPNGPSQQRVIRRALADARLTT
                SDVDVVEAHGTGTRLGDPIEAQAVIATYGQGRDGEQPLRLGSLKSNIGHTQAAAGVSG
                VIKMVQAMRHGVLPKTLHVEKPTDQVDWSAGAVELLTEAMDWPDKGDGGLRRAAVSSF
                GVSGTNAHVVLEEAPAAEETPASEATPAVEPSVGAGLVEPSVLVSAKTPAALDAQIGRLA
                AFASQGRTDAADPGAVARVLAGGRAEFEHRAVVLGTGQDDFAQALTAPEGLIRGTPSD
                VGRVAFVFPGQGTQWAGMGAELLDVSKEFAAAMAECESALSRYVDWSLEAVVRQAPGA
                PTLERVDVVQPVTFAVMVSLAKVWQHHGVTPQAVVGHSQGEIAAAYVAGALTLDDAAR
                VVTLRSKSIAAHLAGKGGMISLALSEEATRQRIENLHGLSIAAVNGPTATVVSGDPTQ
                IQELAQACEADGVRARIIPVDYASHSAHVETIESELAEVLAGLSPRTPEVPFFSTLEG
                AWITEPVLDGTYWYRNLRHRVGFAPAVETLATDEGFTHFIEVSAHPVLTMTLPETVTG
                LGTLRREQGGQERLVTSLAEAWTNGLTIDWAPVLPTATGHHPELPTYAFQRRHYWLHD
                SPAVQGSVQDSWRYRIDWKRLAVADASERAGLSGRWLVVVPEDRSAEAAPVLAALSGA
                GADPVQLDVSPLGDRQRLAATLGEALAAAGGAVDGVLSLLAWDESAHPGHPAPFTRGT
                GATLTLVQALEDAGVAAPLWCVTHGAVSVGRADHVTSPAQAMVWGMGRVAALEHPERW
                GGLIDLPSDADRAALDRMTTVLAGGTGEDQVAVRASGLLARRLVRASLPAHGTASPWW
                QADGTVLVTGAEEPAAAEAARRLARDGAGHLLLHTTPSGSEGAEGTSGAAEDSGLAGL
                VAELADLGATATVVTCDLTDAEAAARLLAGVSDAHPLSAVLHLPPTVDSEPLAATDAD
                ALARVVTAKATAALHLDRLLREAAAAGGRPPVLVLFSSVAAIWGGAGQGAYAAGTAFL
                DALAGQHRADGPTVTSVAWSPWEGSRVTEGATGERLRRLGLRPLAPATALTALDTALG
                HGDTAVTIADVDWSSFAPGFTTARPGTLLADLPEARRALDEQQSTTAADDTVLSRELG
                ALTGAEQQRRMQELVREHLAVVLNHPSPEAVDTGRAFRDLGFDSLTAVELRNRLKNAT
                GLALPATLVFDYPTPRTLAEFLLAEILGEQAGAGEQLPVDGGVDDEPVAIVGMACRLP
                GGVASPEDLWRLVAGGEDAISGFPQDRGWDVEGLYDPDPDASGRTYCRAGGFLDEAGE
```

Query Match 30.5%; Score 133.6; DB 1; Length 37948;
Best Local Similarity 58.2%; Pred. No. 1.1e-15;
Matches 235; Conservative 0; Mismatches 169; Indels 0; Gaps 0;

```
Qy          1 TACTGGCTCGAAAGCACACAGCCCGGTGCCGGCAACGTGTCAGCAGCCGGACTCGACCCC 60
              |||||||   |   |   |   |||  |||||  ||  |   |  | |||||  ||||   |
Db      24073 TACTGGCCGCAGCCCGACCTCTCCGCCGCCGGTGACATCACCTCCGCCGGTCTCGGGGCG 24132

Qy         61 ACCGAACACCCCCTACTCGGCGCCACATTGGAACTGGCGACTGACGGTGGAGCGCTTCTT 120
               ||||  |||||  ||  ||||||||   |  |||   ||  |||      ||  ||     ||  ||
Db      24133 GCCGAGCACCCGCTGCTCGGCGCGGCCGTGGCGCTCGCGGACTCCGACGGCTGCCTGCTC 24192

Qy        121 GCAGGGCGCTTGTCTTTGAGGTCGCATCCGTGGCTGGCTGACCATGCCGTCGGCGGCACG 180
              | |||  ||  |  ||   |   |  ||||  ||  ||||||||  |||||  ||  ||  |  ||||||
Db      24193 ACGGGGAGCCTCTCCCTCCGTACGCACCCCTGGCTGGCGGACCACGCGGTGGCCGGCACC 24252

Qy        181 GTGCTGCTGTCGGGCGCCACCTTCCTCGAACTCGCCCTTCATGCGGGCACATACGTGGGC 240
              |||||||||  ||||   |  |  |||  |  ||  ||  ||   |  |   ||  ||      |  ||  ||
Db      24253 GTGCTGCTGCCGGGAACGGCGTTCGTGGAGCTGGCGTTCCGAGCCGGGGACCAGGTCGGT 24312

Qy        241 TGCGACCGAGTGGATGAGCTGACGCTGCATGCGCCGCTGGTGGTTCCTGTGGATGGGGGT 300
              |||||  |   ||  ||  |||||  ||  ||   |  ||||||||  |||  |  ||       |||  |
Db      24313 TGCGATCTGGTCGAGGAGCTCACCCTCGACGCGCCGCTCGTGCTGCCCCGTCGTGGCGCG 24372

Qy        301 GTGAGTGTGCAGGTTGGGGTTGCGGCTGCGGATGGGGAGGGGCGGCGTTTGGTGAGTGTG 360
              ||  |||||||||  |     ||  |  ||     ||  |  |     |||||  |||     |  |  |
Db      24373 GTCCGTGTGCAGCTGTCCGTCGGCGCGAGCGACGAGTCCGGGCGTCGTACCTTCGGGCTC 24432

Qy        361 TATGCGCGGGGGTGGGAGTGCTTGTGGTGGGGGTGGTGCGTCGGG 404
              ||  ||||     |  |   ||    ||  |  ||   |     |  ||  |
Db      24433 TACGCGCACCCGGAGGACGCGCCGGGCGAGGCGGAGTGGACGCG 24476
```

```
RESULT 13
AX211705
LOCUS       AX211705               65140 bp    DNA     linear   PAT 06-SEP-2001
DEFINITION  Sequence 1 from Patent WO0159126.
ACCESSION   AX211705
VERSION     AX211705.1  GI:15523937
KEYWORDS    .
SOURCE      Streptomyces noursei.
  ORGANISM  Streptomyces noursei
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1  (bases 1 to 65140)
  AUTHORS   Zotchev,S.B., Sekurova,O.N., Fjaervik,E., Brautaset,T.,
            Stroem,A.R., Valla,S., Ellingsen,T.E., Sletta,H.V. and
            Gulliksen,O.M.
  TITLE     Gene cluster encoding a nystatin polyketide synthase and its
            manipulation and utility
  JOURNAL   Patent: WO 0159126-A 1 16-AUG-2001;
            Norges Teknisk Naturvitenskapelige Universitet (NO) ; STIFTELSEN
            IND OG TEKNISK FORSKNING VED NORGES TEKNISKE HOGSKOLE (NO) ;
            ALPHARMA AS (NO) ; Sinvent AS (NO) ; Zotchev, Sergey Borisovich
            (NO) ; Sekurova, Olga Nikolayivna (NO) ; Fjaervik, Espen (NO) ;
            Brautaset, Trygve (NO) ; Strom, Arne Reidar (NO) ; Valla, Svein
            (NO)
FEATURES             Location/Qualifiers
     source          1. .65140
                     /organism="Streptomyces noursei"
                     /db_xref="taxon:1971"
                     /note="ATCC 11455"
BASE COUNT     8270 a  25171 c  22273 g   9426 t
ORIGIN

  Query Match           30.3%;  Score 132.8;  DB 6;  Length 65140;
  Best Local Similarity   60.1%;  Pred. No. 1.3e-15;
  Matches  221;  Conservative    0;  Mismatches  147;  Indels    0;  Gaps    0;

Qy          1 TACTGGCTCGAAAGCACACAGCCCGGTGCCGGCAACGTGTCAGCAGCCGGACTCGACCCC 60
              | ||||| ||    |  ||   || |  ||| | ||||    || ||||| || | | ||
Db      24428 TTCTGGCCCGCCGCCCGACCCGCCCGCCCCGACGACGTCCGCGCCGCCGGCCTGGGCGCC 24487

Qy         61 ACCGAACACCCCCTACTCGGCGCCACATTGGAACTGGCGACTGACGGTGGAGCGCTTCTT 120
              |||||||||||| |||||||| |  || ||||  |    ||||  || || ||  |
Db      24488 GCCGAACACCCCCTCCTCGGCGCCGCCGTGCAACTCCCCGACGACGACGGCGCACTCTTC 24547

Qy        121 GCAGGGCGCTTGTCTTTGAGGTCGCATCCGTGGCTGGCTGACCATGCCGTCGGCGGCACG 180
              | || ||| | ||  || |  | || |||||||||| |||||  |||||   |||||
Db      24548 ACCGGCCGCCTCTCCCTGCGCACCCACCCGTGGCTGGCCGACCACACCGTCCTGGGCACC 24607

Qy        181 GTGCTGCTGTCGGGCGCCACCTTCCTCGAACTCGCCCTTCATGCGGGCACATACGTGGGC 240
              || |||||  ||||| || |  |  | |||||||| | |  ||||||    |    |||
Db      24608 GTCCTGCTCCCGGGCACCGCACTGGTGGAACTCGCCGTCCGCGCGGGCGACGAGACCGGC 24667

Qy        241 TGCGACCGAGTGGATGAGCTGACGCTGCATGCGCCGCTGGTGGTTCCTGTGGATGGGGGT 300
              ||| ||   | || || || || ||     ||||| |||   | || | ||| || |
Db      24668 AGCGGCCACCTCGAAGAACTCACCCTCGCCGCGCCCCTGACCCTCCCCGAGGACGGCGCC 24727

Qy        301 GTGAGTGTGCAGGTTGGGGTTGCGGCTGCGGATGGGGAGGGGCGGCGTTTGGTGAGTGTG 360
                | |||||  | || |   | || || |     || || ||    || |  ||
Db      24728 ACCCTCCTCCAGGTCCGCGTCGGATCCGCCGACGACACCGGCCGCCGCACCGTCACCGTC 24787

Qy        361 TATGCGCG 368
                | || ||
Db      24788 CACGCCCG 24795


RESULT 14
AF263912
LOCUS       AF263912               123580 bp    DNA     linear   BCT 24-MAY-2000
DEFINITION  Streptomyces noursei ATCC 11455 nystatin biosynthetic gene cluster,
            complete sequence.
ACCESSION   AF263912
VERSION     AF263912.1  GI:8050835
KEYWORDS    .
SOURCE      Streptomyces noursei.
  ORGANISM  Streptomyces noursei
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1  (bases 1 to 123580)
  AUTHORS   Brautaset,T., Sekurova,O.N., Sletta,H., Ellingsen,T.E., StrLm,A.R.,
            Valla,S. and Zotchev,S.B.
  TITLE     Biosynthesis of the polyene antifungal antibiotic nystatin in
            Streptomyces noursei ATCC 11455: analysis of the gene cluster and
            deduction of the biosynthetic pathway
  JOURNAL   Chem. Biol. 7 (6), 395-403 (2000)
  MEDLINE   20334850
   PUBMED   10873841
REFERENCE   2  (bases 1 to 123580)
  AUTHORS   Brautaset,T., Sekurova,O.N., Sletta,H., Ellingsen,T.E., Strom,A.R.,
            Valla,S. and Zotchev,S.B.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-MAY-2000) Unigen, NTNU, O. Kyrres gt. 3, Trondheim
            N-7489, Norway
FEATURES             Location/Qualifiers
     source          1. .123580
                     /organism="Streptomyces noursei"
                     /strain="ATCC 11455"
                     /db_xref="ATCC:11455"
                     /db_xref="taxon:1971"
     gene            complement(46. .783)
                     /gene="nysF"
     CDS             complement(46. .783)
                     /gene="nysF"
                     /function="presumably post-translationally modifies ACP
                     domains on PKS"
                     /note="putative 4'-phosphopantheteine transferase"
                     /codon_start=1
                     /transl_table=11
                     /product="NysF"
                     /protein_id="AAF71762.1"
                     /db_xref="GI:8050836"
                     /translation="MIELILPATVATEAAYDDRPRPGDRLLSSEREVIARAVESRQRE
                     FTTVRHLARRALRRLGHPDRAILPNRRGAPQWPPGIVGSMTHCAGYRAAAVSPAELSA
                     AVSIDAEPNGPLPAGVLNAIALPSERPHLVALAAHRPDVHWDRLLFSAKESVFKAWYP
                     LTQRELDFSEAEIVIDPTQGAFTARLLVPGPLLGGRRVTVFPGRWHSTPALLTTAVHL
                     PAPTPRRDREHRTHLTVNSPLPRPTFG"
     gene            complement(867. .2684)
                     /gene="nysG"
     CDS             complement(867. .2684)
                     /gene="nysG"
                     /function="presumably involved in efflux of nystatin"
                     /note="putative transporter (ABC family)"
                     /codon_start=1
                     /transl_table=11
                     /product="NysG"
                     /protein_id="AAF71763.1"
                     /db_xref="GI:8050837"
                     /translation="MASPDDLEEERTAPRPARRLVGLLRPHRRSVALAVSMGVGGIVL
                     NAFGPLLLGRVTDLIADGVLGGVPGPAPGIDFAAIGRLLLVLLALYVVASLFMLAQGR
                     LVASAVWRTIHELRRDAREKLTRLPLRHFDRQPAGELLSRTTNDIDNLQQTLQQTLAE
                     LITSIFSLLTMLVLMLVISPSLAVVMLLSVPVSALIAARISKRAQPHYAAQWSANGTL
                     NAHVEEVCTGHALIKGFDRRAAAEERFDACNDAVYRAAAKAQFASGAMEPVMMFVANL
                     GYVLVAVIGAWKVINGTLTLGDVQAFILYARQFSQPIVEIASVAGRLQSGIASAQRVF
                     TLLDAPEQAPDPLRPGTPARAEGRVEFTDVSFRYSPDTPLIENLSLTVEPGSTVAIVG
                     PTGAGKTTLGNLLMRFYEPDSGRILLDGTDTATMTRDDLRSRFGLVLQDTWLFGGTIA
                     ENIAYGAPGACRADIEEAARATCADRFIRTLPQGYDTVLDDESGTVSAGEKQLLTVAR
                     AFLARPAVLVLDEATSSVDTRTEVLIQRAMNSLRAGRTSFVIAHRLSTIRDADLIVVM
                     DAGRIVEQGTHDQLLCAQGLYARLHAARTHTPTAGAAAG"
     gene            complement(2662. .4416)
                     /gene="nysH"
     CDS             complement(2662. .4416)
                     /gene="nysH"
                     /function="presumably involved in efflux of nystatin"
                     /note="putative transporter (ABC family)"
                     /codon_start=1
```

```
                     /transl_table=11
                     /product="NysH"
                     /protein_id="AAF71764.1"
                     /db_xref="GI:8050838"
                     /translation="MLLRLLRAQLRPYAWATAALVALQLVQILGTLLLPTLGAALIDQ
                     GVVRGDGGRITELGVVMGVVALVQIAAALGAAALAARTATAMGRDLRSALFRRILDFS
                     AREIGRFGTPSLLTRSVNDVQQVQNLAQTGFGIVVCAPLMCLGSVLLALRQDVPLALL
                     LVALVLVVAVCFGLLLARMGTLYARMQLTLDRLGRLLREAITGVRVVRSFVRDDHERA
                     RFAQTNDAFLVVSRRVGRLIATMLPVVLLLMNGFTVALLWTGSHRIDAGRMPIGSLSA
                     LLSYLSLILMSVVMLAFVFLSVPRARVCAGRIAEVLDTGSSVAPPAAPQPVRGPAGRI
                     ELCAAGYRYPGAEEPVLRDVDLTVEPGERIAVLGSTGSGKTTLLNLVLRLADATEGAV
                     RVGGTDVRELTAATLAAAVGFVPQRPYLFSGTVASNLRFGRPDATDEELWEALRVAQA
                     ADFVARMPDGLDAEITQGGGNVSGGQRQRLSLARALLRRPEIYLFDDCFSALDQATDA
                     ALRTALVPYTAGATVITVAQRISAGRDADRIVVLDRGRVVAQGTHDVLLRTSPTYREI
                     ALSQLTEEEAAHGLAGRP"
gene                 4714. .5748
                     /gene="nysDIII"
CDS                  4714. .5748
                     /gene="nysDIII"
                     /function="presumably involved in mycosamine biosynthesis"
                     /note="putative dGDP-mannose-4,6-dehydratase"
                     /codon_start=1
                     /transl_table=11
                     /product="NysDIII"
                     /protein_id="AAF71765.1"
                     /db_xref="GI:8050839"
                     /translation="MSKRALITGITGQDGSYLAEHLLSQGYQVWGLIRGQANPRKSRV
                     SRLASELDFIDGDLMDQGSLVSAVDTVQPDEVYNLGAISFVPMSWQQAELVTEVNGMG
                     VLRMLEAIRMVSGLSTSRTVSPRGQIRFYQASSSEMFGKAAETPQRETTLFHPRSPYG
                     AAKAYGHYITRNYRESFGMYAVSGMLFNHESPRRGQEFVTRKISLAVARIKQGLQDKL
                     ALGNLDAVRDWGYAGDYVRAMHLMLQQDAGDDYVIGTGQMHSVRDAVRIAFEHVGLNW
                     EDYVVIDPDLVRPAEVEVLCADSAKAQDRLGWKPDVDFPTLMRMMVDSDLAQVSRENQ
                     YGDVLLAANW"
gene                 5930. .34363
                     /gene="nysI"
CDS                  5930. .34363
                     /gene="nysI"
                     /function="responsible for condensation steps 9 to 14 in
                     the nystatin polyketide backbone synthesis"
                     /note="polyketide synthase"
                     /codon_start=1
                     /transl_table=11
                     /product="NysI"
                     /protein_id="AAF71766.1"
                     /db_xref="GI:8050840"
                     /translation="MDNEQKLRDYLKLATADLRRTRRRVHKLESAAQEPVAIIGMTCR
                     YPGGVRSPEDLWRMVEAGEHGVTPFPTDRGWDLEALAAAPTASGGFLHDAPDFDADFF
                     GISPREAVAMDPQQRVVLESAWEAFERAGIDPTSVKGSRTGVFIGAMAQDYRVGPADG
                     AEGFQLTGNTGSVLSGRISYTFGTVGPAVTVDTACSSSLVAVHLATQALRAGECTLAL
                     AGGVTIMSGPGTFIEMGRQGGLSADGRCRSFGDTADGTGWAEGVGILVLERLSDAVRN
                     GHEILAVVRGTAVNQDGASNGLTAPNGPSQQQVIQQALVNARLAAGDIDVVEAHGTGT
                     TLGDPVEAQALLATYGQNRPADRPLLLGSVKSNLSHTQAAAGVAGVIKMVMAMRHGTL
                     PRTLHAEEPTHHVDWSQGAVRLLTDTTDWPATGAPRRAAVSSFGISGTNAHTIIEQAP
                     EPQPEDAATAQDDAAGSTPATAPVVPGVVPVLLSGRTPDALRGQAAALRAALDTGRRP
                     DLLDLAHSLATTRAGFEHRAVLLATDHPALTDGLTALADADDPAAAPAWITGTTRAET
                     RLAVLFTGQGAQRLGAGRELAARFPAFATALDAALDAFTPHLDRPLREVLWGTDAALL
                     DRTAYAQPALFAVEVALYRLIESFGVRPDHLAGHSVGEIVAAHLAGVLSLADAATLVA
                     ARGRLMQALPDGGAMIAVQASEADVAPLLAGHEDQVAIAAVNGPSAVVLSGAEATVTA
                     LAEQLAADGRKTRRLRVSHAFHSPLMEPMLDAFRAVVEDLTLQPPLLPVVSNLTGKPA
                     TVAQLTSADYWVDHVRHAVRFADGIDWLARHDTTAFLELGPDGVLSAMAQDCLDAADA
                     DAVTLPALRAGRPEEHTLTTALAGLHVHGATLDWTGCFAGTGARRTDLPTYAFQRRRY
                     WPKALQSGTADLRSVGLGAAHHPLLSAAVSLADAGGTLLTGRLSRQTHPWLADHTVRG
                     TTLLPGTAFLELAVRAGDEVGCDRVEELTLAAPLLLPEQGGVQVQLWIGNPDVSGRRT
                     VNVHARPDTGDDTPWTAHATGVLTTADASRQLPASSEQGGTPLAGDPHPALDAAQWPP
                     AGAEPLPLDGHYDRLADGGFGYGPVFQGLRAAWRGGDVVYAEVELPEAGRSDAEAFGL
                     HPALLDAALHAAPFTGLGERGRGGLPFSWEGVSLHAGGATTLRVRLTPVADDALALTV
                     ADGTGAPVLSVDSLVLRSVATQQLDTAAAVARDALFRLDWTPVQPTATDPGPVALLGA
                     DPFGLLTHAGFADAPAYPDLAALAAADGPVPTTVVLSLAGTGDDAADPARSAHRCAAE
                     ALAAVQTWLDHHERFAAARLVFVTRGATVGRDVAAAAVWGLVRSAQSENPGCFALVDL
                     DPDGAVGAAALVAALVSGEPQLAVRGDVLRVARLVRRPLTEVGAGADGTGDGVGDGSG
                     VSFSGEGAVLVTGGTGGLGAVHARHLVAEYGVRDLLLVSRSGERAVGAGELVAELAGV
                     GARVRVVACDVTDRAAVVELVGGHAVSAVVHAAGVLDDGMVGALTGERLSAVLRPKVD
                     AVWHLHEATRGLDLDAFVVFSSLAGVFGSPGQANYAAANAFLDALMTRRRAEGLPGLS
                     LAWGPWEQSGGMTGTLTDVDAERLARSGVPPLSVAQGLALFDAAVAGTDATCVPVRLD
                     LPVLRARGEVPPLLRSLIRVRARRAAVAGSATAGNLAQRLRRLDEDGRDEMVLDLVRG
                     QVALVLGHATGGDVDAGRAFRDLGFDSLTAVELRNRLNTVTGLRLPATLVFDYPTVRH
                     LATYVLDELLGTDAEVATVQPAAVAVADDPIVIVGMACRYPGGVSSPEDLWRVLTEGT
                     DAVSGFPTNRGWDVESLYHPDPDHPGTSYTRSGGFLHEAGEFDPGFFGMSPREALATD
                     SQQRLLLESSWEAIERAGIDPVSLRGSRTGVFAGVMYSDYSAMLASPEPEGFQGSGSS
                     PSLASGRVAYTLGLEGPAVTVDTACSSSLVAMHWAMQALRSGECGLALAGGVTVMSTP
                     AVFVDFARQRGLSPDGRCKAFADAADGVGWSEGVGVLVLERQSDAVRNGHEILAVVRG
                     SAVNQDGASNGLTAPNGPSQQRVIRQALASGGLTAGDVDVVEAHGTGTTLGDPIEAQA
                     LLATYGRDREPERPLLLGSVKSNLGHTQAAAGVAGVIKMVLAMRHGVVPRTLHVDAPS
                     SHVDWSEGAVELLSEQAAWPETGRVRRAGVSSFGISGTNVHVIVEQAPGAKAIAAAGA
                     ARRTPGAVPVLLSGRGRSALRGQAARLLGHLQARPDAELVDVALSLATTRSRFEQRAA
                     VVAQDRDQLIASLGALAADRPDPAVVEGEAAGRGRTAVLFTGQGSQRAAMGRELHEVQ
                     PEFAAAFDAVCAVFDPLLDRPLREVVFAEDGSDEAALLDETGWTQPALFAVEVALFRL
                     VESWGVRPDFVAGHSIGEIAAAHVAGVLTLEDACRLVAARATLMQALPTGGAMIAIQA
                     TEDEIAAHLDDTVAIAAVNGPQSVVISGDEEAAETIAATFAERGRKTKRLRVSHAFHS
                     PRMDGMLDAFRIVAEGLTYRAPRIPLVSDLTGRRADDAEVCTAEYWVRHVREAVRFAD
                     CVRTLRDAGATTFLELGSDGLLTAMAEDTLGDDHDAELVPMLRAGRAEELAAATALAR
                     LQVRGVDVDWAAYLAGTGARRTDLPTYAFQHAYYWPQLPTPAAALAAADPADQQLWAA
                     VERGDARELADILGLGEQDLTPLDSLLPALTSWRRGNQEKHLLDTLRYRVEWTRLSKP
                     TAPVLDGTWLLVASDATAADQPALLDGLADALGSHGARVRRLLLDDSCADRAVLAERL
                     ARTADVDAATQVLSVLPLDERDADDCPPLTRGLALTVALVQALADTGAQGRLWTATRG
                     AVSTNPADPVTHPVQAAAWGLGRGVALEHPRLWGGLVDLPQVFDERAGQRLAGILAVK
                     DAPDGEDQVALRATGVSGRRLVRHTVEALPTAAEFTATGTVLITGGTGGLGAEVARWL
                     ARAGAQHLVLTSRRGPDAPGAAELRAELEGYGPSVSVVACDVADRDALAAVLTALPEE
                     LPLTGVVHTAGVGHYGPLDTLSTAEFAGLTAAKLAGAAHLDALLADRELDFFVLFGSI
                     AGVWGSGNQSAYGAANAYLDALALHRRARGLAATSVAWGPWAEAGMAADDAVSETLRR
                     QGLGLLDPAPAMTELRRAVVRQDVTVTVADVDWQRYAPLFTSARPSALIAGLPEVRAL
                     AADERTEQDATGASEVVTRVRALAEPEQLRLLTDLVRTESATVLGHSSADAVPEGRAF
                     RDVGFDSLTAVELRKRLGAATGLSLPSTMVFDYPTPLELAQYLRAEILGAVLEVAGPV
                     ATGGADDEPIAIIGMACRFPGGVSSPEQLWDLVASGTDAISEFPVNRGWQTGHLFDPD
                     PDRPGTTYSTQGGFLHEADEFDPTFFGISPREALVMDPQQRLLLETTWESFERAGIRP
                     ETLRSTLTGTFVGSSYQEYGLGAGDGTEGHMVTGSSPSVLSGRLSYVFGLEGPAVTVD
                     TACSSSLVALHLACQSLRNGESNLAVAGGATIMTTPNPFIAFSRQRALAKDGRCKAFS
                     DDADGMTLAEGVGVVLVERLSDAQRNGHPVLAVLRGSAINQDGASNGLTAPNGPSQQR
                     VIRQALANARLAPGDIDALEAHGTGTPLGDPIEAQALFATYGRDRDPESALLLGSVKS
                     NIGHTQSAAGIASVIKMVMALRHSELPPTLHADAPSSHVDWSAGTVRLLTQARAWPET
                     GRPRRAAVSSFGISGTNAHVLLEQAPVADTPAEERPAVAPVPIAAGVVPWVVTARSAA
                     ALRGQAERLLAHAETVGTALPAAGPLDIGLSLVSARARFEHRAVVVPPAGTDPLAALR
                     AVATDGPSPVVARGVADVEGRTVFVFPGQGSQWVGMGSQLLDESAVFAERIAECAAAL
                     AEFTDWSLVDVLRGVVGAPSLERVDVVQPASFAVMVSLAALWRSRGVLPDAVVGHSQG
                     EIAAAVVSGALSLRDGARVVALRSQAIGRALAGRGGMMSVALSVDVLEPRLVEFEGRV
                     SVAAVNGPRSVVVAGEPEALDALHARLTADDIRARRIAVDYASHSHQVEDLHEELLEV
                     LAELAPRTSEVPFFSTVTGDWLDTARMDAGYWFRNLRGRVRFADAVADLLAAEYRAFV
                     EVSSHPVLSMAVQEAIDEAGVPAVAAGTLRRDQGGTDRFLLSAAEVFVRGVDVDWAGL
                     FEGTGASRIDLPTYAFQHEHLWAVPPAPEAVAAADPDDAAFWTAVEDGDVSALTAALG
                     TDEDSVAAVLPALTSWRRARRDRSTVDAWRYRVAWKPLGGTLPHPSLTGTWLLVTADG
                     IDDTDVAGALETYGAEVRRLVLDEECVDRAVLRERLAGAEDVTGIVSVLAAAERTDAV
                     PGTSLVLGTALTVALIQALGDAEIDAPVWALTRGAVSTGRADELTAPVQAQVTGIGWT
                     AALEHPQRWGGTLDLPAALDARAAQRLAAVLSGALGSDDQLAIRPSGVFTRRIVRAEA
```

Query Match 30.3%; Score 132.8; DB 1; Length 123580;
Best Local Similarity 60.1%; Pred. No. 1.2e-15;
Matches 221; Conservative 0; Mismatches 147; Indels 0; Gaps 0;

```
Qy       1 TACTGGCTCGAAAGCACACAGCCCGGTGCCGGCAACGTGTCAGCAGCCGGACTCGACCCC 60
           | ||||| ||    |  ||   || |  ||| | ||||    || ||||| || | | ||
Db   84281 TTCTGGCCCGCCGCCCGACCCGCCCGCCCCGACGACGTCCGCGCCGCCGGCCTGGGCGCC 84340

Qy      61 ACCGAACACCCCCTACTCGGCGCCACATTGGAACTGGCGACTGACGGTGGAGCGCTTCTT 120
           ||||||||||||| |||||||| |  || ||||  |    ||||  || || ||  |
Db   84341 GCCGAACACCCCCTCCTCGGCGCCGCCGTGCAACTCCCCGACGACGACGGCGCACTCTTC 84400

Qy     121 GCAGGGCGCTTGTCTTTGAGGTCGCATCCGTGGCTGGCTGACCATGCCGTCGGCGGCACG 180
           | || ||| | ||  || |  | || |||||||||| |||||  |||||  |||||
Db   84401 ACCGGCCGCCTCTCCCTGCGCACCCACCCGTGGCTGGCCGACCACACCGTCCTGGGCACC 84460

Qy     181 GTGCTGCTGTCGGGCGCCACCTTCCTCGAACTCGCCCTTCATGCGGGCACATACGTGGGC 240
           || |||||  ||||| || |  |  | |||||||||.| | |  ||||||    |   |||
Db   84461 GTCCTGCTCCCGGGCACCGCACTGGTGGAACTCGCCGTCCGCGCGGGCGACGAGACCGGC 84520

Qy     241 TGCGACCGAGTGGATGAGCTGACGCTGCATGCGCCGCTGGTGGTTCCTGTGGATGGGGGT 300
           ||| ||   | || || || || ||    ||||| |||    | || | ||| || |
```

```
Db      84521 AGCGGCCACCTCGAAGAACTCACCCTCGCCGCGCCCCTGACCCTCCCCGAGGACGGCGCC 84580

Qy        301 GTGAGTGTGCAGGTTGGGGTTGCGGCTGCGGATGGGGAGGGGCGGCGTTTGGTGAGTGTG 360
              | |||||  | || |   | || || |     || || ||    || |  ||
Db      84581 ACCCTCCTCCAGGTCCGCGTCGGATCCGCCGACGACACCGGCCGCCGCACCGTCACCGTC 84640

Qy        361 TATGCGCG 368
              | || ||
Db      84641 CACGCCCG 84648


RESULT 15
AX211739
LOCUS       AX211739              125401 bp    DNA     linear   PAT 06-SEP-2001
DEFINITION  Sequence 35 from Patent WO0159126.
ACCESSION   AX211739
VERSION     AX211739.1  GI:15523950
KEYWORDS    .
SOURCE      Streptomyces noursei.
  ORGANISM  Streptomyces noursei
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1  (bases 1 to 125401)
  AUTHORS   Zotchev,S.B., Sekurova,O.N., Fjaervik,E., Brautaset,T.,
            Stroem,A.R., Valla,S., Ellingsen,T.E., Sletta,H.V. and
            Gulliksen,O.M.
  TITLE     Gene cluster encoding a nystatin polyketide synthase and its
            manipulation and utility
  JOURNAL   Patent: WO 0159126-A 35 16-AUG-2001;
            Norges Teknisk Naturvitenskapelige Universitet (NO) ; STIFTELSEN
            IND OG TEKNISK FORSKNING VED NORGES TEKNISKE HOGSKOLE (NO) ;
            ALPHARMA AS (NO) ; Sinvent AS (NO) ; Zotchev, Sergey Borisovich
            (NO) ; Sekurova, Olga Nikolayivna (NO) ; Fjaervik, Espen (NO) ;
            Brautaset, Trygve (NO) ; Strom, Arne Reidar (NO) ; Valla, Svein
            (NO)
FEATURES             Location/Qualifiers
     source          1. .125401
                     /organism="Streptomyces noursei"
                     /db_xref="taxon:1971"
                     /note="ATCC 11455"
BASE COUNT    15664 a  49692 c  42871 g  17174 t
ORIGIN

  Query Match           30.3%;  Score 132.8;  DB 6;  Length 125401;
  Best Local Similarity   60.1%;  Pred. No. 1.2e-15;
  Matches  221;  Conservative    0;  Mismatches  147;  Indels    0;  Gaps    0;

Qy          1 TACTGGCTCGAAAGCACACAGCCCGGTGCCGGCAACGTGTCAGCAGCCGGACTCGACCCC 60
              | ||||| ||    |  ||   || |  ||| | ||||    || ||||| || | | ||
Db      84689 TTCTGGCCCGCCGCCCGACCCGCCCGCCCCGACGACGTCCGCGCCGCCGGCCTGGGCGCC 84748

Qy         61 ACCGAACACCCCCTACTCGGCGCCACATTGGAACTGGCGACTGACGGTGGAGCGCTTCTT 120
              |||||||||||||| |||||||| |  || ||||  |    ||||  || || ||  |
Db      84749 GCCGAACACCCCCTCCTCGGCGCCGCCGTGCAACTCCCCGACGACGACGGCGCACTCTTC 84808

Qy        121 GCAGGGCGCTTGTCTTTGAGGTCGCATCCGTGGCTGGCTGACCATGCCGTCGGCGGCACG 180
              | || ||| | ||  || |  | || |||||||||| |||||  |||||   |||||
Db      84809 ACCGGCCGCCTCTCCCTGCGCACCCACCCGTGGCTGGCCGACCACACCGTCCTGGGCACC 84868

Qy        181 GTGCTGCTGTCGGGCGCCACCTTCCTCGAACTCGCCCTTCATGCGGGCACATACGTGGGC 240
              || |||||  ||||| || |  |  | |||||||||| | |  ||||||    |    |||
Db      84869 GTCCTGCTCCCGGGCACCGCACTGGTGGAACTCGCCGTCCGCGCGGGCGACGAGACCGGC 84928

Qy        241 TGCGACCGAGTGGATGAGCTGACGCTGCATGCGCCGCTGGTGGTTCCTGTGGATGGGGGT 300
              ||| ||   | || || || || ||    ||||| |||    | || | ||| || |
Db      84929 AGCGGCCACCTCGAAGAACTCACCCTCGCCGCGCCCCTGACCCTCCCCGAGGACGGCGCC 84988

Qy        301 GTGAGTGTGCAGGTTGGGGTTGCGGCTGCGGATGGGGAGGGGCGGCGTTTGGTGAGTGTG 360
              | |||||  | || |   | || || |     || || ||    || |  ||
Db      84989 ACCCTCCTCCAGGTCCGCGTCGGATCCGCCGACGACACCGGCCGCCGCACCGTCACCGTC 85048

Qy        361 TATGCGCG 368
              | || ||
Db      85049 CACGCCCG 85056


Search completed: June 17, 2003, 17:55:49
Job time : 852.086 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 17, 2003, 11:45:13 ; Search time 18.9737 Seconds
(without alignments)
7079.508 Million cell updates/sec

Title: US-09-914-286-1_COPY_8947_9384
Perfect score: 438
Sequence: 1 tactggctcgaaagcacaca.....gtcatgcctcgggggtgctg 438

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	133.6	30.5	11220	4	US-09-105-537-32	Sequence 32, Appl
2	133.6	30.5	36778	4	US-09-105-537-5	Sequence 5, Appli
3	133.6	30.5	38506	3	US-09-320-878-19	Sequence 19, Appl
c 4	124	28.3	1590	4	US-09-434-288-1	Sequence 1, Appli
5	123.4	28.2	80161	3	US-09-036-987A-1	Sequence 1, Appli
6	123.4	28.2	80161	4	US-09-370-700-1	Sequence 1, Appli
7	122.2	27.9	43280	2	US-08-804-227C-1	Sequence 1, Appli
8	117.4	26.8	50937	4	US-09-428-517-1	Sequence 1, Appli
9	111.8	25.5	13987	2	US-08-804-227C-13	Sequence 13, Appl
10	111.8	25.5	44377	2	US-08-804-227C-7	Sequence 7, Appli
11	111.8	25.5	44377	2	US-08-804-198-1	Sequence 1, Appli
12	109.4	25.0	33529	4	US-09-144-085-3	Sequence 3, Appli
c 13	105.8	24.2	4403765	4	US-09-103-840A-2	Sequence 2, Appli
c 14	105.8	24.2	4411529	4	US-09-103-840A-1	Sequence 1, Appli
15	105.6	24.1	49377	1	US-08-764-233A-1	Sequence 1, Appli
16	105.2	24.0	15872	4	US-09-105-537-1	Sequence 1, Appli
17	104.8	23.9	13842	4	US-09-105-537-30	Sequence 30, Appl
18	101.2	23.1	784	1	US-08-392-731-1	Sequence 1, Appli
19	99	22.6	28958	1	US-08-258-261B-6	Sequence 6, Appli
20	99	22.6	28958	1	US-08-456-837-6	Sequence 6, Appli
21	99	22.6	28958	1	US-08-457-342-6	Sequence 6, Appli
22	99	22.6	28958	1	US-08-457-646A-6	Sequence 6, Appli
23	99	22.6	28958	1	US-08-458-076A-6	Sequence 6, Appli
24	99	22.6	28958	1	US-08-764-233A-4	Sequence 4, Appli
25	99	22.6	28958	1	US-08-457-335A-6	Sequence 6, Appli
26	99	22.6	28958	1	US-08-729-214-6	Sequence 6, Appli
27	99	22.6	28958	3	US-09-028-934-6	Sequence 6, Appli
28	95	21.7	20235	1	US-07-642-734C-3	Sequence 3, Appli
29	95	21.7	20235	3	US-08-439-009A-3	Sequence 3, Appli
30	92.4	21.1	4411529	4	US-09-103-840A-1	Sequence 1, Appli
31	83	18.9	4403765	4	US-09-103-840A-2	Sequence 2, Appli
32	52.4	12.0	320	4	US-09-165-264-14	Sequence 14, Appl
33	52.2	11.9	320	4	US-09-165-264-13	Sequence 13, Appl
34	51.2	11.7	320	4	US-09-165-264-7	Sequence 7, Appli
35	50.4	11.5	318	4	US-09-165-264-12	Sequence 12, Appl
36	50.4	11.5	319	4	US-09-165-264-8	Sequence 8, Appli
37	50.4	11.5	320	4	US-09-165-264-11	Sequence 11, Appl
c 38	48.6	11.1	152331	3	US-09-128-155-16	Sequence 16, Appl
c 39	46	10.5	12001	1	US-08-458-568A-11	Sequence 11, Appl
c 40	42.8	9.8	2830	1	US-07-882-292-1	Sequence 1, Appli
c 41	42.8	9.8	2830	2	US-08-331-644-1	Sequence 1, Appli
c 42	42.8	9.8	2830	5	PCT-US93-04102-1	Sequence 1, Appli
c 43	42.2	9.6	521	4	US-09-643-597-324	Sequence 324, App
c 44	41.6	9.5	697	4	US-09-040-984-17	Sequence 17, Appl
c 45	41.6	9.5	697	4	US-09-123-912-17	Sequence 17, Appl

ALIGNMENTS

RESULT 1
US-09-105-537-32
; Sequence 32, Application US/09105537A
; Patent No. 6265202
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/105,537A
; CURRENT FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 32
; LENGTH: 11220
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-105-537-32

Query Match 30.5%; Score 133.6; DB 4; Length 11220;
Best Local Similarity 58.2%; Pred. No. 6.9e-23;
Matches 235; Conservative 0; Mismatches 169; Indels 0; Gaps 0;

```
Qy          1 TACTGGCTCGAAAGCACACAGCCCGGTGCCGGCAACGTGTCAGCAGCCGGACTCGACCCC 60
              ||||||| |   |   |  ||| ||||| || |  | | ||||| ||||  |
Db       7216 TACTGGCCGCAGCCCGACCTCTCCGCCGCCGGTGACATCACCTCCGCCGGTCTCGGGGCG 7275

Qy         61 ACCGAACACCCCCTACTCGGCGCCACATTGGAACTGGCGACTGACGGTGGAGCGCTTCTT 120
               |||| ||||| || |||||||| |  |||  || |||     ||  ||    || ||
Db       7276 GCCGAGCACCCGCTGCTCGGCGCGGCCGTGGCGCTCGCGGACTCCGACGGCTGCCTGCTC 7335

Qy        121 GCAGGGCGCTTGTCTTTGAGGTCGCATCCGTGGCTGGCTGACCATGCCGTCGGCGGCACG 180
              | ||| || | ||  |  |  |||| || |||||||| ||||| || || | ||||||
Db       7336 ACGGGGAGCCTCTCCCTCCGTACGCACCCCTGGCTGGCGGACCACGCGGTGGCCGGCACC 7395

Qy        181 GTGCTGCTGTCGGGCGCCACCTTCCTCGAACTCGCCCTTCATGCGGGCACATACGTGGGC 240
              |||||||| |||| |  | ||| | || || ||  | |  || ||      | || ||
Db       7396 GTGCTGCTGCCGGGAACGGCGTTCGTGGAGCTGGCGTTCCGAGCCGGGGACCAGGTCGGT 7455

Qy        241 TGCGACCGAGTGGATGAGCTGACGCTGCATGCGCCGCTGGTGGTTCCTGTGGATGGGGGT 300
              ||||| |  || || ||||| || ||  | |||||||| ||| | ||       ||| |
Db       7456 TGCGATCTGGTCGAGGAGCTCACCCTCGACGCGCCGCTCGTGCTGCCCCGTCGTGGCGCG 7515

Qy        301 GTGAGTGTGCAGGTTGGGGTTGCGGCTGCGGATGGGGAGGGGCGGCGTTTGGTGAGTGTG 360
              ||  |||||||| |    || |  ||    || | |   ||||| |||    |  |  |
Db       7516 GTCCGTGTGCAGCTGTCCGTCGGCGCGAGCGACGAGTCCGGGCGTCGTACCTTCGGGCTC 7575

Qy        361 TATGCGCGGGGTGGGAGTGCTTGTGGTGGGGGTGGTGCGTCGGG 404
```

```
Db       7576 TACGCGCACCCGGAGGACGCGCCGGGCGAGGCGGAGTGGACGCG 7619


RESULT 2
US-09-105-537-5
; Sequence 5, Application US/09105537A
; Patent No. 6265202
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT: Liu, H.
;  APPLICANT: Xue, Y.
;  APPLICANT: Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/105,537A
;  CURRENT FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 5
;   LENGTH: 36778
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-105-537-5

  Query Match             30.5%;  Score 133.6;  DB 4;  Length 36778;
  Best Local Similarity   58.2%;  Pred. No. 8.2e-23;
  Matches  235;  Conservative    0;  Mismatches  169;  Indels    0;  Gaps    0;

Qy          1 TACTGGCTCGAAAGCACACAGCCCGGTGCCGGCAACGTGTCAGCAGCCGGACTCGACCCC 60
              ||||||| |  |  |  ||| ||||| || |  | | ||||| |||| |
Db      22903 TACTGGCCGCAGCCCGACCTCTCCGCCGCCGGTGACATCACCTCCGCCGGTCTCGGGGCG 22962

Qy         61 ACCGAACACCCCCTACTCGGCGCCACATTGGAACTGGCGACTGACGGTGGAGCGCTTCTT 120
              |||| ||||| || ||||||||  |  |||  || |||     ||  ||    || ||
Db      22963 GCCGAGCACCCGCTGCTCGGCGCGGCCGTGGCGCTCGCGGACTCCGACGGCTGCCTGCTC 23022

Qy        121 GCAGGGCGCTTGTCTTTGAGGTCGCATCCGTGGCTGGCTGACCATGCCGTCGGCGGCACG 180
              | ||| || | ||  |  |  |||| || |||||||| ||||| || || | ||||||
Db      23023 ACGGGGAGCCTCTCCCTCCGTACGCACCCCTGGCTGGCGGACCACGCGGTGGCCGGCACC 23082

Qy        181 GTGCTGCTGTCGGGCGCCACCTTCCTCGAACTCGCCCTTCATGCGGGCACATACGTGGGC 240
              ||||||||| ||||  |  | ||| | || || ||  | |  || ||      | || ||
Db      23083 GTGCTGCTGCCGGGAACGGCGTTCGTGGAGCTGGCGTTCCGAGCCGGGGACCAGGTCGGT 23142

Qy        241 TGCGACCGAGTGGATGAGCTGACGCTGCATGCGCCGCTGGTGGTTCCTGTGGATGGGGGT 300
              |||||  |  || || ||||| || ||  | |||||||| ||| | ||      ||| |
Db      23143 TGCGATCTGGTCGAGGAGCTCACCCTCGACGCGCCGCTCGTGCTGCCCCGTCGTGGCGCG 23202

Qy        301 GTGAGTGTGCAGGTTGGGGTTGCGGCTGCGGATGGGGAGGGGCGGCGTTTGGTGAGTGTG 360
              ||  |||||||| |    || | ||    || | |   ||||| |||    |  | |
Db      23203 GTCCGTGTGCAGCTGTCCGTCGGCGCGAGCGACGAGTCCGGGCGTCGTACCTTCGGGCTC 23262

Qy        361 TATGCGCGGGGTGGGAGTGCTTGTGGTGGGGGTGGTGCGTCGGG 404
              || ||||    | |  ||    || | ||  |    | || |
Db      23263 TACGCGCACCCGGAGGACGCGCCGGGCGAGGCGGAGTGGACGCG 23306


RESULT 3
US-09-320-878-19
; Sequence 19, Application US/09320878A
; Patent No. 6117659
; GENERAL INFORMATION:
;  APPLICANT: ASHLEY, Gary
;  APPLICANT: BETLACH, Melanie C.
;  APPLICANT: BETLACH, Mary C.
;  APPLICANT: McDANIEL, Robert
;  APPLICANT: TANG, Li
;  TITLE OF INVENTION: RECOMBINANT NARBONOLIDE POLYKETIDE SYNTHASE
;  FILE REFERENCE: 300622002120
;  CURRENT APPLICATION NUMBER: US/09/320,878A
;  CURRENT FILING DATE: 1999-05-27
;  EARLIER APPLICATION NUMBER: CIP OF 09/141,908
;  EARLIER FILING DATE: 1998-08-28
;  EARLIER APPLICATION NUMBER: CIP OF 09/073,538
;  EARLIER FILING DATE: 1998-05-06
;  EARLIER APPLICATION NUMBER: CIP OF 08/846,247
;  EARLIER FILING DATE: 1997-04-30
;  EARLIER APPLICATION NUMBER: 60/119,139
;  EARLIER FILING DATE: 1999-02-08
;  EARLIER APPLICATION NUMBER: 60/100,880
;  EARLIER FILING DATE: 1998-09-22
;  EARLIER APPLICATION NUMBER: 60/087,080
;  EARLIER FILING DATE: 1998-05-28
;  NUMBER OF SEQ ID NOS: 34
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 19
;   LENGTH: 38506
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-320-878-19

  Query Match             30.5%;  Score 133.6;  DB 3;  Length 38506;
  Best Local Similarity   58.2%;  Pred. No. 8.3e-23;
  Matches  235;  Conservative    0;  Mismatches  169;  Indels    0;  Gaps    0;

Qy          1 TACTGGCTCGAAAGCACACAGCCCGGTGCCGGCAACGTGTCAGCAGCCGGACTCGACCCC 60
              |||||||  |  |  |  ||| |||||  || |  | | ||||| |||| |
Db      21045 TACTGGCCGCAGCCCGACCTCTCCGCCGCCGGTGACATCACCTCCGCCGGTCTCGGGGCG 21104

Qy         61 ACCGAACACCCCCTACTCGGCGCCACATTGGAACTGGCGACTGACGGTGGAGCGCTTCTT 120
              |||| ||||| || ||||||||  |  |||  || |||     ||  ||    || ||
Db      21105 GCCGAGCACCCGCTGCTCGGCGCGGCCGTGGCGCTCGCGGACTCCGACGGCTGCCTGCTC 21164

Qy        121 GCAGGGCGCTTGTCTTTGAGGTCGCATCCGTGGCTGGCTGACCATGCCGTCGGCGGCACG 180
              | ||| || | ||  |  |  |||| || |||||||| ||||| || || | ||||||
Db      21165 ACGGGGAGCCTCTCCCTCCGTACGCACCCCTGGCTGGCGGACCACGCGGTGGCCGGCACC 21224

Qy        181 GTGCTGCTGTCGGGCGCCACCTTCCTCGAACTCGCCCTTCATGCGGGCACATACGTGGGC 240
              |||||||| ||||  |  | ||| | || || ||  | |  || ||      | || ||
Db      21225 GTGCTGCTGCCGGGGAACGGCGTTCGTGGAGCTGGCGTTCCGAGCCGGGGACCAGGTCGGT 21284

Qy        241 TGCGACCGAGTGGATGAGCTGACGCTGCATGCGCCGCTGGTGGTTCCTGTGGATGGGGGT 300
              |||||  |  || || ||||| || ||  | |||||||| ||| | ||      ||| |
Db      21285 TGCGATCTGGTCGAGGAGCTCACCCTCGACGCGCCGCTCGTGCTGCCCCGTCGTGGCGCG 21344

Qy        301 GTGAGTGTGCAGGTTGGGGTTGCGGCTGCGGATGGGGAGGGGCGGCGTTTGGTGAGTGTG 360
              ||  |||||||| |    || | ||    || | |   ||||| |||    |  | |
Db      21345 GTCCGTGTGCAGCTGTCCGTCGGCGCGAGCGACGAGTCCGGGCGTCGTACCTTCGGGCTC 21404

Qy        361 TATGCGCGGGGTGGGAGTGCTTGTGGTGGGGGTGGTGCGTCGGG 404
              || ||||    | |  ||    || | ||  |    | || |
Db      21405 TACGCGCACCCGGAGGACGCGCCGGGCGAGGCGGAGTGGACGCG 21448


RESULT 4
US-09-434-288-1/c
; Sequence 1, Application US/09434288
; Patent No. 6303767
; GENERAL INFORMATION:
;  APPLICANT:  Betlach C., Melanie
;  APPLICANT:  McDaniel, Robert
;  TITLE OF INVENTION:  POLYKETIDE SYNTHASE ENZYMES AND RECOMBINANT DNA
;  TITLE OF INVENTION:  CONSTRUCTS THEREFOR
;  FILE REFERENCE:  30062-20030.00
;  CURRENT APPLICATION NUMBER: US/09/434,288
;  CURRENT FILING DATE:  1999-11-05
;  PRIOR APPLICATION NUMBER:  60/107,093
;  PRIOR FILING DATE:  1998-11-05
;  NUMBER OF SEQ ID NOS:  12
;  SOFTWARE:  PatentIn Ver. 2.1
; SEQ ID NO 1
;   LENGTH: 1590
;   TYPE: DNA
;   ORGANISM: Streptomyces narbonensis
US-09-434-288-1
```

```
  Query Match             28.3%;  Score 124;  DB 4;  Length 1590;
  Best Local Similarity   59.8%;  Pred. No. 9.1e-21;
  Matches  208;  Conservative    0;  Mismatches  140;  Indels    0;  Gaps    0;

Qy          1 TACTGGCTCGAAAGCACACAGCCCGGTGCCGGCAACGTGTCAGCAGCCGGACTCGACCCC 60
              ||||||| |    ||  |||  | |||  || |  | | ||||| ||||    |
Db        752 TACTGGCCGCGGCCCGAGCACTCCGCCGTCGGTGACATCACGTCCGCCGGTCTCGGGGCG 693

Qy         61 ACCGAACACCCCCTACTCGGCGCCACATTGGAACTGGCGACTGACGGTGGAGCGCTTCTT 120
              |||| ||||| || |||||||| |  | |  || ||      ||  ||    || ||
Db        692 GCCGAGCACCCGCTGCTCGGCGCGGCCGTCGCGCTCGCCGACTCCGACGGCTGCCTGCTC 633

Qy        121 GCAGGGCGCTTGTCTTTGAGGTCGCATCCGTGGCTGGCTGACCATGCCGTCGGCGGCACG 180
              | ||| || | ||  |  |  |||| || |||||||| ||||| || || | |||||||
Db        632 ACGGGGAGCCTCTCCCTCCGTACGCACCCCTGGCTGGCGGACCACGCGGTGGCCGGCACG 573

Qy        181 GTGCTGCTGTCGGGCGCCACCTTCCTCGAACTCGCCCTTCATGCGGGCACATACGTGGGC 240
              ||||||||| ||||  |  | ||| | || || ||| | |  || ||      | || |||
Db        572 GTGCTGCTGCCGGGAACGGCGTTCGTGGAGCTGGCCTTCCGGGCCGGTGACCAGGTCGGC 513

Qy        241 TGCGACCGAGTGGATGAGCTGACGCTGCATGCGCCGCTGGTGGTTCCTGTGGATGGGGGT 300
              |||||  |  || || ||||| || ||  | || ||||| ||  | ||      || |
Db        512 TGCGATCTGGTCGAGGAGCTCACCCTCGACGCCCCGCTCGTCCTGCCCCGTCGCGGCGCG 453

Qy        301 GTGAGTGTGCAGGTTGGGGTTGCGGCTGCGGATGGGGAGGGGCGGCGT 348
              ||  | ||||||  |    || |  ||    || | |   ||||| |||
Db        452 GTCCGCGTGCAGCTGTCCGTCGGCGCGAGCGACGAGTCCGGGCGCCGT 405


RESULT 5
US-09-036-987A-1
; Sequence 1, Application US/09036987A
; Patent No. 6143526
;  GENERAL INFORMATION:
;    APPLICANT:  Baltz, Richard H.
;    APPLICANT:  Broughton, Mary C.
;    APPLICANT:  Crawford, Kathryn P.
;    APPLICANT:  Madduri, Krishnamurthy
;    APPLICANT:  Merlo, Donald J.
;    APPLICANT:  Treadway, Patti J.
;    APPLICANT:  Turner, Jan R.
;    APPLICANT:  Waldron, Clive
;    TITLE OF INVENTION:  Biosynthetic Genes For Spinosyn Insecticide
;    TITLE OF INVENTION:  Production
;    NUMBER OF SEQUENCES:  39
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Dow AgroSciences LLC Patent Department
;      STREET:  9330 Zionsville Road
;      CITY:  Indianapolis
;      STATE:  Indiana
;      COUNTRY:  USA
;      ZIP:  46268
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/036,987A
;      FILING DATE:  09-MAR-1998
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Stuart, Donald R
;      REGISTRATION NUMBER:  28,479
;      REFERENCE/DOCKET NUMBER:  50,608
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (317)337-4816
;      TELEFAX:  (317)337-4847
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  80161 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-09-036-987A-1

  Query Match             28.2%;  Score 123.4;  DB 3;  Length 80161;
  Best Local Similarity   58.2%;  Pred. No. 2.2e-20;
  Matches  217;  Conservative    0;  Mismatches  156;  Indels    0;  Gaps    0;

Qy          1 TACTGGCTCGAAAGCACACAGCCCGGTGCCGGCAACGTGTCAGCAGCCGGACTCGACCCC 60
              |||||| | ||   | |  ||  ||  || ||| |  |      | |  ||  |   | ||
Db      62518 TACTGGTTGGAGTCCGCCGAGTACGACGCGGGCGATCTCGGTTCGGTGGGCTTGCTCTCC 62577

Qy         61 ACCGAACACCCCCTACTCGGCGCCACATTGGAACTGGCGACTGACGGTGGAGCGCTTCTT 120
              |||| || ||||| |||||  ||  |  ||   |||||   ||  || ||     || ||
Db      62578 GCCGAGCATCCCCTGCTCGGGGCTGCGGTGACGCTGGCCGATGCGGGCGGGTTCCTGCTG 62637

Qy        121 GCAGGGCGCTTGTCTTTGAGGTCGCATCCGTGGCTGGCTGACCATGCCGTCGGCGGCACG 180
              | ||      |||| | | | | || || ||| |||| |||||  |  ||||||||  ||
Db      62638 ACCGGCAAGCTGTCGGTCAAGACCCAGCCCTGGTTGGCCGACCACGTGGTCGGCGGGGCG 62697

Qy        181 GTGCTGCTGTCGGGCGCCACCTTCCTCGAACTCGCCCTTCATGCGGGCACATACGTGGGC 240
              | ||||||  | ||| || | ||| | ||| |      | |  || |      | || ||
Db      62698 ATCCTGCTGCCCGGCACCGCGTTCGTGGAAATGCTGATACGCGCCGCGGACCAGGTCGGG 62757

Qy        241 TGCGACCGAGTGGATGAGCTGACGCTGCATGCGCCGCTGGTGGTTCCTGTGGATGGGGGT 300
              |||||  |    | || ||| || | |||    | ||||||||  | || | |   || |
Db      62758 TGCGATCTGATCGAGGAGTTGTCCCTGACGACTCCGCTGGTTTTGCCCGCGACCGGTGCG 62817

Qy        301 GTGAGTGTGCAGGTTGGGGTTGCGGCTGCGGATGGGGAGGGGCGGCGTTTGGTGAGTGTG 360
              |||   ||||||  | | |||||  | | |||| | ||  |||||  || | |||  | |||
Db      62818 GTGCAGGTGCAGATCGCGGTTGGCGGTCCGGACGAGGCCGGGCGCCGCTCGGTCCGCGTG 62877

Qy        361 TATGCGCGGGGTG 373
              ||| |  |  | |
Db      62878 CATTCCTGTCGAG 62890


RESULT 6
US-09-370-700-1
; Sequence 1, Application US/09370700
; Patent No. 6274350
; GENERAL INFORMATION:
;  APPLICANT: Baltz, Richard H
;  APPLICANT: Broughton, Mary C
;  APPLICANT: Crawford, Kathryn P
;  APPLICANT: Madduri, Krishnamurthy
;  APPLICANT: Treadway, Patti J
;  APPLICANT: Turner, Jan R
;  APPLICANT: Waldron, Clive
;  TITLE OF INVENTION: Biosynthetic Genes For Spinosyn Insecticide
;  FILE REFERENCE: 50489 DIV1
;  CURRENT APPLICATION NUMBER: US/09/370,700
;  CURRENT FILING DATE: 1999-08-09
;  EARLIER APPLICATION NUMBER: US 09/36987
;  EARLIER FILING DATE: 1998-03-09
;  NUMBER OF SEQ ID NOS: 39
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
;   LENGTH: 80161
;   TYPE: DNA
;   ORGANISM: Saccharopolyspora spinosa
US-09-370-700-1

  Query Match             28.2%;  Score 123.4;  DB 4;  Length 80161;
  Best Local Similarity   58.2%;  Pred. No. 2.2e-20;
  Matches  217;  Conservative    0;  Mismatches  156;  Indels    0;  Gaps    0;

Qy          1 TACTGGCTCGAAAGCACACAGCCCGGTGCCGGCAACGTGTCAGCAGCCGGACTCGACCCC 60
              |||||| | ||   | |  ||  ||  || ||| |  |      | |  ||  |   | ||
Db      62518 TACTGGTTGGAGTCCGCCGAGTACGACGCGGGCGATCTCGGTTCGGTGGGCTTGCTCTCC 62577

Qy         61 ACCGAACACCCCCTACTCGGCGCCACATTGGAACTGGCGACTGACGGTGGAGCGCTTCTT 120
```

```
Db      62578 GCCGAGCATCCCCTGCTCGGGGCTGCGGTGACGCTGGCCGATGCGGGCGGGTTCCTGCTG 62637

Qy        121 GCAGGGCGCTTGTCTTTGAGGTCGCATCCGTGGCTGGCTGACCATGCCGTCGGCGGCACG 180

Db      62638 ACCGGCAAGCTGTCGGTCAAGACCCAGCCCTGGTTGGCCGACCACGTGGTCGGCGGGGCG 62697

Qy        181 GTGCTGCTGTCGGGCGCCACCTTCCTCGAACTCGCCCTTCATGCGGGCACATACGTGGGC 240

Db      62698 ATCCTGCTGCCCGGCACCGCGTTCGTGGAAATGCTGATACGCGCCGCGGACCAGGTCGGG 62757

Qy        241 TGCGACCGAGTGGATGAGCTGACGCTGCATGCGCCGCTGGTGGTTCCTGTGGATGGGGGT 300

Db      62758 TGCGATCTGATCGAGGAGTTGTCCCTGACGACTCCGCTGGTTTTGCCCGCGACCGGTGCG 62817

Qy        301 GTGAGTGTGCAGGTTGGGGTTGCGGCTGCGGATGGGGAGGGGCGGCGTTTGGTGAGTGTG 360

Db      62818 GTGCAGGTGCAGATCGCGGTTGGCGGTCCGGACGAGGCCGGGCGCCGCTCGGTCCGCGTG 62877

Qy        361 TATGCGCGGGGTG 373

Db      62878 CATTCCTGTCGAG 62890


RESULT 7
US-08-804-227C-1
; Sequence 1, Application US/08804227C
; Patent No. 5876991
;  GENERAL INFORMATION:
;    APPLICANT:  DeHoff, Bradley S.
;    APPLICANT:  Kuhstoss, Stuart A.
;    APPLICANT:  Rosteck, Paul R., Jr.
;    APPLICANT:  Sutton, Kimberly L.
;    TITLE OF INVENTION:  POLYKETIDE SYNTHASE GENES
;    NUMBER OF SEQUENCES:  15
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  THOMAS G. PLANT  1501
;      STREET:  LILLY CORPORATE CENTER
;      CITY:  INDIANAPOLIS
;      STATE:  IN
;      COUNTRY:  USA
;      ZIP:  46285
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  MS-DOS
;      SOFTWARE:  ASCI(DOS) Text only
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/804,227C
;      FILING DATE:  February 21, 1997
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Plant, Thomas, G.
;      REGISTRATION NUMBER:  35,784
;      REFERENCE/DOCKET NUMBER:  X-8231
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  317-276-2459
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  43280 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  816..14234
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  14351..19945
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  20010..31199
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  31232..36067
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  36249..41774
US-08-804-227C-1

  Query Match             27.9%;  Score 122.2;  DB 2;  Length 43280;
  Best Local Similarity   57.0%;  Pred. No. 3.9e-20;
  Matches  223;  Conservative    0;  Mismatches  168;  Indels    0;  Gaps    0;

Qy          1 TACTGGCTCGAAAGCACACAGCCCGGTGCCGGCAACGTGTCAGCAGCCGGACTCGACCCC 60

Db      27330 TACTGGTACCACCCCACGAGCGGCAGGCGCGGCGACATCACGGCGGCCGGCGTGGCCGAG 27389

Qy         61 ACCGAACACCCCCTACTCGGCGCCACATTGGAACTGGCGACTGACGGTGGAGCGCTTCTT 120

Db      27390 GCGGAGCACCCGCTGCTCGGCGCCGGCGTCGAACTCCCGGAGACCGGGGGCACGGTGTAC 27449

Qy        121 GCAGGGCGCTTGTCTTTGAGGTCGCATCCGTGGCTGGCTGACCATGCCGTCGGCGGCACG 180

Db      27450 ACCGCGCGGTTCGGCCCGGACAGCCGGCCGTGGCTGGCCGACCACGCGCTGCTGGGCACC 27509

Qy        181 GTGCTGCTGTCGGGCGCCACCTTCCTCGAACTCGCCCTTCATGCGGGCACATACGTGGGC 240

Db      27510 GTGCTGCTGCCCGGCACGGCAATCCTGGACCTGGTGCTGTGGGCGGGCGAACGCTCCGGC 27569

Qy        241 TGCGACCGAGTGGATGAGCTGACGCTGCATGCGCCGCTGGTGGTTCCTGTGGATGGGGGT 300

Db      27570 TGCGGCCGCGTCGGTGAACTGGCGCTCCAGGCACCGCTGGTCCTGCCGGACAGCGGCGAC 27629

Qy        301 GTGAGTGTGCAGGTTGGGGTTGCGGCTGCGGATGGGGAGGGGCGGCGTTTGGTGAGTGTG 360

Db      27630 GTCGAACTGCGGCTGCTGGTCGGCGGCCCGGACGAGGAGAAACGGCGCACCGTCACCGTG 27689

Qy        361 TATGCGCGGGGTGGGAGTGCTTGTGGTGGGG 391

Db      27690 CACGCGCGGCCCGCGGCGGCCGGCGCCGAGG 27720


RESULT 8
US-09-428-517-1
; Sequence 1, Application US/09428517
; Patent No. 6251636
; GENERAL INFORMATION:
;  APPLICANT: Betlach, Mary C.
;  APPLICANT: Shah, Sanjay Krishnakant
;  APPLICANT: McDaniel, Robert
;  APPLICANT: Tang, Li
;  TITLE OF INVENTION: RECOMBINANT OLEANDOLIDE POLYKETIDE SYNTHASE
;  FILE REFERENCE: 30062-20029.00
;  CURRENT APPLICATION NUMBER: US/09/428,517
;  CURRENT FILING DATE: 1999-10-28
;  EARLIER APPLICATION NUMBER: 60/120,254
;  EARLIER FILING DATE: 1999-02-16
;  EARLIER APPLICATION NUMBER: 60/106,100
;  EARLIER FILING DATE: 1998-10-29
;  NUMBER OF SEQ ID NOS: 12
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
;   LENGTH: 50937
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artifical Sequence: Recombinant DNA
US-09-428-517-1

  Query Match             26.8%;  Score 117.4;  DB 4;  Length 50937;
  Best Local Similarity   60.5%;  Pred. No. 5.3e-19;
  Matches  193;  Conservative    0;  Mismatches  126;  Indels    0;  Gaps    0;

Qy         31 GGCAACGTGTCAGCAGCCGGACTCGACCCCACCGAACACCCCCTACTCGGCGCCACATTG 90
```

```
Db      25716 GGCGACGTGGCGTCCGTGGGCCTCACCGCGGCCGGCCACCCCCTTCTGGGCGCGGTCGTG 25775

Qy         91 GAACTGGCGACTGACGGTGGAGCGCTTCTTGCAGGGCGCTTGTCTTTGAGGTCGCATCCG 150
              ||  ||  |       ||  ||   | | ||  | ||||   | ||  ||  || | | |||||
Db      25776 GAGATGCCCGACTCCGACGGGTTGGTGCTCACCGGGCAGATCTCCCTGCGGACCCATCCC 25835

Qy        151 TGGCTGGCTGACCATGCCGTCGGCGGCACGGTGCTGCTGTCGGGCGCCACCTTCCTCGAA 210
              |||||  ||  |||||  |  ||    |||  |||||||  |||  |||||  ||  |  |||  ||||
Db      25836 TGGCTCGCCGACCACGAGGTGCTCGGATCGGTGCTCCTGCCGGGCACCGCGTTCGTCGAG 25895

Qy        211 CTCGCCCTTCATGCGGGCACATACGTGGGCTGCGACCGAGTGGATGAGCTGACGCTGCAT 270
              ||  |||  |  ||  ||  |  |       ||  ||  |  ||||   |  ||||  ||||||||||||  |
Db      25896 CTTGCCGTCCAGGCCGCCGACCGCGCCGGTTACGACGTACTGGACGAGCTGACGCTGGAG 25955

Qy        271 GCGCCGCTGGTGGTTCCTGTGGATGGGGGTGTGAGTGTGCAGGTTGGGGTTGCGGCTGCG 330
              |||||  ||  |||  |  ||  |       ||  ||   |    ||||   |  |    |  |  |  |
Db      25956 GCGCCCCTCGTGCTCCCCGACAGGGGCGGCATCCAGGTGCGTCTGGCCCTCGGGCCGTCC 26015

Qy        331 GATGGGGAGGGGCGGCGTT 349
              ||  |   ||  ||  ||  ||  |
Db      26016 GAGGCAGACGGACGCCGGT 26034


RESULT 9
US-08-804-227C-13
; Sequence 13, Application US/08804227C
; Patent No. 5876991
;  GENERAL INFORMATION:
;    APPLICANT:  DeHoff, Bradley S.
;    APPLICANT:  Kuhstoss, Stuart A.
;    APPLICANT:  Rosteck, Paul R., Jr.
;    APPLICANT:  Sutton, Kimberly L.
;    TITLE OF INVENTION:  POLYKETIDE SYNTHASE GENES
;    NUMBER OF SEQUENCES:  15
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  THOMAS G. PLANT  1501
;      STREET:  LILLY CORPORATE CENTER
;      CITY:  INDIANAPOLIS
;      STATE:  IN
;      COUNTRY:  USA
;      ZIP:  46285
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  MS-DOS
;      SOFTWARE:  ASCI(DOS) Text only
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/804,227C
;      FILING DATE:  February 21, 1997
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Plant, Thomas, G.
;      REGISTRATION NUMBER:  35,784
;      REFERENCE/DOCKET NUMBER:  X-8231
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  317-276-2459
;  INFORMATION FOR SEQ ID NO:  13:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  13987 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  350..13987
US-08-804-227C-13

  Query Match             25.5%;  Score 111.8;  DB 2;  Length 13987;
  Best Local Similarity   57.6%;  Pred. No. 9e-18;
  Matches  200;  Conservative    0;  Mismatches  147;  Indels    0;  Gaps    0;

Qy         48 CGGACTCGACCCCACCGAACACCCCCTACTCGGCGCCACATTGGAACTGGCGACTGACGG 107
              |||  ||  ||     |  || ||||||  |  |||  |||  |  |    |    | ||       |||
Db      10987 CGGCCTGGAGTGGAAGGACCACCCCTTCCTCAGCGGCGCCACGCCGATAGCCGGCTCCGG 11046

Qy        108 TGGAGCGCTTCTTGCAGGGCGCTTGTCTTTGAGGTCGCATCCGTGGCTGGCTGACCATGC 167
              |     |||  ||  |  ||  |  ||      |      | ||  ||||||||||| |||||  |
Db      11047 CGCGCTGCTCCTCACCGGCAGGGTGGGGCTCGCTGCCCACCCGTGGCTGGCCGACCACGC 11106

Qy        168 CGTCGGCGGCACGGTGCTGCTGTCGGGCGCCACCTTCCTCGAACTCGCCCTTCATGCGGG 227
              | ||  |||||||||||||||  |  ||  |  ||  |||  ||     ||  |  ||||
Db      11107 CATCTCCGGCACGGTGCTGCTCCCCGGAACGGCGATCGCCGACCTGCTGCTGCGGGCGGT 11166

Qy        228 CACATACGTGGGCTGCGACCGAGTGGATGAGCTGACGCTGCATGCGCCGCTGGTGGTTCC 287
              |     | ||  |||  ||    | ||  ||  ||  ||  ||||||||  ||||  |||  |||  |   | ||
Db      11167 CGAGGAGGTCGGCGCCGGAGGGGTCGAGGAACTGACGCTCCATGAGCCCCTGCTCCTCCC 11226

Qy        288 TGTGGATGGGGGTGTGAGTGTGCAGGTTGGGGTTGCGGCTGCGGATGGGGAGGGGCGGCG 347
              | |    ||  ||   ||    ||  |||||    |||  |  |||  ||  ||  |  |  ||||  |||||
Db      11227 CGAGCGAGGCGGCCTGCACGTCCAGGTGCTGGTCGAGGCGGCCGACGAGCAGGGACGGCG 11286

Qy        348 TTTGGTGAGTGTGTATGCGCGGGGTGGGAGTGCTTGTGGTGGGGGTG 394
              |   |||    ||    ||  ||    |  |  |  ||  |   |  ||||
Db      11287 TGCCGTGGCAGTCGCCGCACGCCCGGAGGGCCCTGGGCGGGACGGTG 11333


RESULT 10
US-08-804-227C-7
; Sequence 7, Application US/08804227C
; Patent No. 5876991
;  GENERAL INFORMATION:
;    APPLICANT:  DeHoff, Bradley S.
;    APPLICANT:  Kuhstoss, Stuart A.
;    APPLICANT:  Rosteck, Paul R., Jr.
;    APPLICANT:  Sutton, Kimberly L.
;    TITLE OF INVENTION:  POLYKETIDE SYNTHASE GENES
;    NUMBER OF SEQUENCES:  15
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  THOMAS G. PLANT  1501
;      STREET:  LILLY CORPORATE CENTER
;      CITY:  INDIANAPOLIS
;      STATE:  IN
;      COUNTRY:  USA
;      ZIP:  46285
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  MS-DOS
;      SOFTWARE:  ASCI(DOS) Text only
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/804,227C
;      FILING DATE:  February 21, 1997
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Plant, Thomas, G.
;      REGISTRATION NUMBER:  35,784
;      REFERENCE/DOCKET NUMBER:  X-8231
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  317-276-2459
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  44377 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  350..14002
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  14046..20036
;    FEATURE:
;      NAME/KEY:  CDS
```

```
;      LOCATION:  20110..31284
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  31329..36071
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  36155..41830
US-08-804-227C-7

  Query Match             25.5%;  Score 111.8;  DB 2;  Length 44377;
  Best Local Similarity   57.6%;  Pred. No. 1.1e-17;
  Matches  200;  Conservative    0;  Mismatches  147;  Indels    0;  Gaps    0;

Qy         48 CGGACTCGACCCCACCGAACACCCCCTACTCGGCGCCACATTGGAACTGGCGACTGACGG 107
              ||| || ||    |  || ||||||| | ||| ||| | |   |    | ||      |||
Db      11002 CGGCCTGGAGTGGAAGGACCACCCCTTCCTCAGCGGCGCCACGCCGATAGCCGGCTCCGG 11061

Qy        108 TGGAGCGCTTCTTGCAGGGCGCTTGTCTTTGAGGTCGCATCCGTGGCTGGCTGACCATGC 167
                |    ||| ||  | ||  |  ||    |    | || ||||||||||| ||||| ||
Db      11062 CGCGCTGCTCCTCACCGGCAGGGTGGGGCTCGCTGCCCACCCGTGGCTGGCCGACCACGC 11121

Qy        168 CGTCGGCGGCACGGTGCTGCTGTCGGGCGCCACCTTCCTCGAACTCGCCCTTCATGCGGG 227
              | ||  |||||||||||||||  | ||  |  |  ||  ||| ||    || |  ||||
Db      11122 CATCTCCGGCACGGTGCTGCTCCCCGGAACGGCGATCGCCGACCTGCTGCTGCGGGCGGT 11181

Qy        228 CACATACGTGGGCTGCGACCGAGTGGATGAGCTGACGCTGCATGCGCCGCTGGTGGTTCC 287
              |    | || |||  ||   | || || || |||||||| |||| ||| ||| |  | ||
Db      11182 CGAGGAGGTCGGCGCCGGAGGGGTCGAGGAACTGACGCTCCATGAGCCCCTGCTCCTCCC 11241

Qy        288 TGTGGATGGGGGTGTGAGTGTGCAGGTTGGGGTTGCGGCTGCGGATGGGGAGGGGCGGCG 347
              | |   || ||  ||   || |||||  ||| | ||| || || | | |||| |||||
Db      11242 CGAGCGAGGCGGCCTGCACGTCCAGGTGCTGGTCGAGGCGGCCGACGAGCAGGGACGGCG 11301

Qy        348 TTTGGTGAGTGTGTATGCGCGGGGTGGGAGTGCTTGTGGTGGGGGTG 394
              |  |||   ||    || ||    | | |  || |  | |  ||||
Db      11302 TGCCGTGGCAGTCGCCGCACGCCCGGAGGGCCCTGGGCGGGACGGTG 11348


RESULT 11
US-08-804-198-1
; Sequence 1, Application US/08804198
; Patent No. 5945320
;  GENERAL INFORMATION:
;    APPLICANT:  Burgett, Stanley G.
;    APPLICANT:  Kuhstoss, Stuart A.
;    APPLICANT:  Rao, Nagaraja R.
;    APPLICANT:  Richardson, Mark A.
;    APPLICANT:  Rosteck, Paul R., Jr.
;    TITLE OF INVENTION:  PLATENOLIDE SYNTHASE GENE
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  PAUL R. CANTRELL  1138
;      STREET:  LILLY CORPORATE CENTER
;      CITY:  INDIANAPOLIS
;      STATE:  IN
;      COUNTRY:  USA
;      ZIP:  46285
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  Macintosh
;      OPERATING SYSTEM:  Macintosh 7.0
;      SOFTWARE:  Microsoft Word 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/804,198
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  CANTRELL, PAUL R.
;      REGISTRATION NUMBER:  36,470
;      REFERENCE/DOCKET NUMBER:  P9113
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  317-276-3885
;  INFORMATION FOR SEQ ID NO:  1:
```

```
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  44377 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  350..14002
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  14046..20036
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  20110..31284
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  31329..36071
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  36155..41830
US-08-804-198-1

  Query Match             25.5%;  Score 111.8;  DB 2;  Length 44377;
  Best Local Similarity   57.6%;  Pred. No. 1.1e-17;
  Matches  200;  Conservative    0;  Mismatches  147;  Indels    0;  Gaps    0;

Qy         48 CGGACTCGACCCCACCGAACACCCCCTACTCGGCGCCACATTGGAACTGGCGACTGACGG 107
              ||| || ||    |  || ||||||| | ||| ||| | |   |    | ||      |||
Db      11002 CGGCCTGGAGTGGAAGGACCACCCCTTCCTCAGCGGCGCCACGCCGATAGCCGGCTCCGG 11061

Qy        108 TGGAGCGCTTCTTGCAGGGCGCTTGTCTTTGAGGTCGCATCCGTGGCTGGCTGACCATGC 167
                |    ||| ||  | ||  |  ||    |    | || ||||||||||| ||||| ||
Db      11062 CGCGCTGCTCCTCACCGGCAGGGTGGGGCTCGCTGCCCACCCGTGGCTGGCCGACCACGC 11121

Qy        168 CGTCGGCGGCACGGTGCTGCTGTCGGGCGCCACCTTCCTCGAACTCGCCCTTCATGCGGG 227
              | ||  |||||||||||||||  | ||  |  |  ||  ||| ||    || |  ||||
Db      11122 CATCTCCGGCACGGTGCTGCTCCCCGGAACGGCGATCGCCGACCTGCTGCTGCGGGCGGT 11181

Qy        228 CACATACGTGGGCTGCGACCGAGTGGATGAGCTGACGCTGCATGCGCCGCTGGTGGTTCC 287
              |    | || |||  ||   | || || || |||||||| |||| ||| ||| |  | ||
Db      11182 CGAGGAGGTCGGCGCCGGAGGGGTCGAGGAACTGACGCTCCATGAGCCCCTGCTCCTCCC 11241

Qy        288 TGTGGATGGGGGTGTGAGTGTGCAGGTTGGGGTTGCGGCTGCGGATGGGGAGGGGCGGCG 347
              | |   || ||  ||   || |||||  ||| | ||| || || | | |||| |||||
Db      11242 CGAGCGAGGCGGCCTGCACGTCCAGGTGCTGGTCGAGGCGGCCGACGAGCAGGGACGGCG 11301

Qy        348 TTTGGTGAGTGTGTATGCGCGGGGTGGGAGTGCTTGTGGTGGGGGTG 394
              |  |||   ||    || ||    | | |  || |  | |  ||||
Db      11302 TGCCGTGGCAGTCGCCGCACGCCCGGAGGGCCCTGGGCGGGACGGTG 11348


RESULT 12
US-09-144-085-3
; Sequence 3, Application US/09144085
; Patent No. 6280999
; GENERAL INFORMATION:
;  APPLICANT: Gustafsson, Claes
;  APPLICANT: Betlach, Mary C.
;  APPLICANT: Ashley, Gary
;  APPLICANT: Julien, Bryan
;  APPLICANT: Ziermann, Rainer
;  TITLE OF INVENTION: SORANGIUM POLYKETIDE SYNTHASES AND ENCODING DNA
;  TITLE OF INVENTION: THEREFOR
;  FILE REFERENCE: 30062-20020.20
;  CURRENT APPLICATION NUMBER: US/09/144,085
;  CURRENT FILING DATE: 1998-08-31
;  EARLIER APPLICATION NUMBER: 09/010,809
;  EARLIER FILING DATE: 1998-01-22
;  NUMBER OF SEQ ID NOS: 8
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
;   LENGTH: 33529
```

```
;   TYPE: DNA
;   ORGANISM: Sorangium cellulosum
US-09-144-085-3

  Query Match             25.0%;  Score 109.4;  DB 4;  Length 33529;
  Best Local Similarity   57.0%;  Pred. No. 3.7e-17;
  Matches  200;  Conservative    0;  Mismatches  151;  Indels    0;  Gaps    0;

Qy         30 CGGCAACGTGTCAGCAGCCGGACTCGACCCCACCGAACACCCCCTACTCGGCGCCACATT 89
              ||| |||||    | || || ||  |  |  ||| || || || |||||||||  |  .
Db       2417 CGGTGACGTGAGCTCGGCGGGGCTGAAGGCGGCCGCCCATCCGCTGCTCGGCGCCGCGAC 2476

Qy         90 GGAACTGGCGACTGACGGTGGAGCGCTTCTTGCAGGGCGCTTGTCTTTGAGGTCGCATCC 149
              | | || ||    | ||  ||    ||  |  ||||| |  ||||  ||  |  |||| |
Db       2477 GAAGCTCGCCGACGGCGAGGGGCACCTGTTCACAGGGAGGCTGTCGCTGGCGGAGCATGC 2536

Qy        150 GTGGCTGGCTGACCATGCCGTCGGCGGCACGGTGCTGCTGTCGGGCGCCACCTTCCTCGA 209
              ||||||    || ||||  ||    |||  |||| || |  |||||  |    | | || ||
Db       2537 GTGGCTTCGGGATCATGGTGTGTTTGGCCAGGTGGTGTTTCCGGGCACGGGCATGCTGGA 2596

Qy        210 ACTCGCCCTTCATGCGGGCACATACGTGGGCTGCGACCGAGTGGATGAGCTGACGCTGCA 269
              | || ||    || ||        |||||| ||        ||   |||||  |||||
Db       2597 GGTTGCGCTGGCGGCTGGGCGCGCGGTGGGCAGCCGGTCGCTGTCGGAGCTCACGCTCGC 2656

Qy        270 TGCGCCGCTGGTGGTTCCTGTGGATGGGGGTGTGAGTGTGCAGGTTGGGGTTGCGGCTGC 329
              | ||||||||||| |  | | ||| || |  | | |  |||||||    | | |  ||  |
Db       2657 CGAGCCGCTGGTGCTGGCCGAGGACGGCGCGGCGCGGCTGCAGGTGATGATCGGAGCGCC 2716

Qy        330 GGATGGGGAGGGGCGGCGTTTGGTGAGTGTGTATGCGCGGGGTGGGAGTGC 380
              ||||  || ||| |||||    ||||  |  ||||     |  |  ||  |  |||
Db       2717 AGATGCGGCGGGCCGGCGCGAGGTGGGGCTGTACAGCCAGCCTGAGCATGC 2767


RESULT 13
US-09-103-840A-2/c
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
;  APPLICANT: FLEISCHMAN, Robert D.
;  APPLICANT:  WHITE, Owen R.
;  APPLICANT:  FRASER, Claire M.
;  APPLICANT:  VENTER, John C.
;  TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
;  TITLE OF INVENTION:  TUBERCULOSIS
;  FILE REFERENCE: 24366-20007.00
;  CURRENT APPLICATION NUMBER: US/09/103,840A
;  CURRENT FILING DATE:  1998-06-24
;  NUMBER OF SEQ ID NOS: 2
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
;   LENGTH: 4403765
;   TYPE: DNA
;   ORGANISM: Mycobacterium tuberculosis
;   FEATURE:
;   OTHER INFORMATION: CDC 1551
;   OTHER INFORMATION: "n" bases at various positions throughout the sequence
;   OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

  Query Match             24.2%;  Score 105.8;  DB 4;  Length 4403765;
  Best Local Similarity   56.4%;  Pred. No. 5.3e-16;
  Matches  197;  Conservative    0;  Mismatches  152;  Indels    0;  Gaps    0;

Qy         25 GGTGCCGGCAACGTGTCAGCAGCCGGACTCGACCCCACCGAACACCCCCTACTCGGCGCC 84
              ||  ||| | | | | | |      || || | | | ||||| ||  ||  | || ||
Db    3289336 GGGCCCGCCGATGCGGCCGGGTTGGGTCTGGGCGCGACCGAGCATGCCTTGTTGGGTGCG 3289277

Qy         85 ACATTGGAACTGGCGACTGACGGTGGAGCGCTTCTTGCAGGGCGCTTGTCTTTGAGGTCG 144
                 | || | | |   |  | | | | | ||  | || || |||||  |   |
Db    3289276 GTGGTCGAGCGGCCCGATTCTGACGAGGTGGTGCTGACCGGCCGGTTGTCGCTTGCGGAT 3289217

Qy        145 CATCCGTGGCTGGCTGACCATGCCGTCGGCGGCACGGTGCTGCTGTCGGGCGCCACCTTC 204
              || |||||||||| ||||| |  ||   |||   ||||||| |  | || ||    ||
Db    3289216 CAGCCGTGGCTGGCCGACCACGTGGTGAACGGGGTGGTGCTGTTCCCCGGGGCGGGTTTT 3289157

Qy        205 CTCGAACTCGCCCTTCATGCGGGCACATACGTGGGCTGCGACCGAGTGGATGAGCTGACG 264
              | ||| | |   | | | || ||      | || || |||| |   | || ||| ||  |
Db    3289156 GTGGAGTTGGTGATCCGCGCCGGTGATGAGGTCGGGTGCGCGCTCATCGAAGAGTTGGTG 3289097

Qy        265 CTGCATGCGCCGCTGGTGGTTCCTGTGGATGGGGGTGTGAGTGTGCAGGTTGGGGTTGCG 324
              |||   || ||| |||| | |    || ||  || ||    |||||||| |  || | |
Db    3289096 CTGGCCGCACCGTTGGTGATGCACCCGGGTGTCGGGGTTCAGGTGCAGGTGGTCGTCGGG 3289037

Qy        325 GCTGCGGATGGGGAGGGGCGGCGTTTGGTGAGTGTGTATGCGCGGGGTG 373
              ||||| ||||     ||||  |||  ||||   |||||| | || ||||
Db    3289036 GCTGCCGATGAATCCGGGCACCGTGCGGTGTCGGTGTATTCCCGCGGTG 3288988


RESULT 14
US-09-103-840A-1/c
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
;  APPLICANT: FLEISCHMAN, Robert D.
;  APPLICANT:  WHITE, Owen R.
;  APPLICANT:  FRASER, Claire M.
;  APPLICANT:  VENTER, John C.
;  TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
;  TITLE OF INVENTION:  TUBERCULOSIS
;  FILE REFERENCE: 24366-20007.00
;  CURRENT APPLICATION NUMBER: US/09/103,840A
;  CURRENT FILING DATE:  1998-06-24
;  NUMBER OF SEQ ID NOS: 2
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
;   LENGTH: 4411529
;   TYPE: DNA
;   ORGANISM: Mycobacterium tuberculosis
;   OTHER INFORMATION: H37Rv
US-09-103-840A-1

  Query Match             24.2%;  Score 105.8;  DB 4;  Length 4411529;
  Best Local Similarity   56.4%;  Pred. No. 5.3e-16;
  Matches  197;  Conservative    0;  Mismatches  152;  Indels    0;  Gaps    0;

Qy         25 GGTGCCGGCAACGTGTCAGCAGCCGGACTCGACCCCACCGAACACCCCCTACTCGGCGCC 84
              ||  ||| | | | | | |      || || | | | ||||| ||  ||  | || ||
Db    3295060 GGGCCCGCCGATGCGGCCGGGTTGGGTCTGGGCGCGACCGAGCATGCCTTGTTGGGTGCG 3295001

Qy         85 ACATTGGAACTGGCGACTGACGGTGGAGCGCTTCTTGCAGGGCGCTTGTCTTTGAGGTCG 144
                 | || | | |   |  | | | | | ||  | || || |||||  |   |
Db    3295000 GTGGTCGAGCGGCCCGATTCTGACGAGGTGGTGCTGACCGGCCGGTTGTCGCTTGCGGAT 3294941

Qy        145 CATCCGTGGCTGGCTGACCATGCCGTCGGCGGCACGGTGCTGCTGTCGGGCGCCACCTTC 204
              || |||||||||| ||||| |  ||   |||   ||||||| |  | || ||    ||
Db    3294940 CAGCCGTGGCTGGCCGACCACGTGGTGAACGGGGTGGTGCTGTTCCCCGGGGCGGGTTTT 3294881

Qy        205 CTCGAACTCGCCCTTCATGCGGGCACATACGTGGGCTGCGACCGAGTGGATGAGCTGACG 264
              | ||| | |   | | | || ||      | || || |||| |   | || ||| ||  |
Db    3294880 GTGGAGTTGGTGATCCGCGCCGGTGATGAGGTCGGGTGCGCGCTCATCGAAGAGTTGGTG 3294821

Qy        265 CTGCATGCGCCGCTGGTGGTTCCTGTGGATGGGGGTGTGAGTGTGCAGGTTGGGGTTGCG 324
              |||   || ||| |||| | |    || ||  || ||    |||||||| |  || | |
Db    3294820 CTGGCCGCACCGTTGGTGATGCACCCGGGTGTCGGGGTTCAGGTGCAGGTGGTCGTCGGG 3294761

Qy        325 GCTGCGGATGGGGAGGGGCGGCGTTTGGTGAGTGTGTATGCGCGGGGTG 373
              ||||| ||||     ||||  |||  ||||   |||||| | || ||||
Db    3294760 GCTGCCGATGAATCCGGGCACCGTGCGGTGTCGGTGTATTCCCGCGGTG 3294712


RESULT 15
US-08-764-233A-1
; Sequence 1, Application US/08764233A
; Patent No. 5716849
;  GENERAL INFORMATION:
;    APPLICANT:  Ligon, James M.
```

```
;    APPLICANT:  Schupp, Thomas
;    APPLICANT:  Beck, James J.
;    APPLICANT:  Hill, Dwight S.
;    APPLICANT:  Neff, Snezanna
;    APPLICANT:  Ryals, John A.
;    TITLE OF INVENTION:  Genes For The Biosynthesis Of Soraphen
;    NUMBER OF SEQUENCES:  10
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Ciba-Geigy Corporation
;      STREET:  520 White Plains Road, P.O. Box 2005
;      CITY:  Tarrytown
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10591
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/764,233A
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/729,214
;      FILING DATE:  09-OCT-1996
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/258,261
;      FILING DATE:  08-JUN-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Meigs, J. Timothy
;      REGISTRATION NUMBER:  38,241
;      REFERENCE/DOCKET NUMBER:  1506/CIP6
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (919) 541-8587
;      TELEFAX:  (919) 541-8689
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  49377 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    ORIGINAL SOURCE:
;      ORGANISM:  Sorangium cellulosum
;    IMMEDIATE SOURCE:
;      CLONE:  p98/1, pJL3, and pVKM15
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  383..760
;      OTHER INFORMATION:  /product= "SorR"
;      OTHER INFORMATION:  /note= "This gene encodes a protein that is highly homologous
;      OTHER INFORMATION:  the reductase domains of type I PKSs such as eryA from
;      OTHER INFORMATION:  Saccharopolyspora erythraea."
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  927..19874
;      OTHER INFORMATION:  /product= "SorA"
;      OTHER INFORMATION:  /note= "Gene product is highly homologous to type I PKSs that
;      OTHER INFORMATION:  are known to be involved in the synthesis of polyketide
;      OTHER INFORMATION:  compounds."
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  942..7115
;      OTHER INFORMATION:  /product= "Module 1 of SorA"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  7203..12884
;      OTHER INFORMATION:  /product= "Module 2 of SorA"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  13455..19616
;      OTHER INFORMATION:  /product= "Module 3 of SorA"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  19871..46318
;      OTHER INFORMATION:  /product= "SorB"
;      OTHER INFORMATION:  /note= "Gene product is highly homologous to type I PKS gene.
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  19870..24556
;      OTHER INFORMATION:  /product= "Module 1 of SorB"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  24638..30820
;      OTHER INFORMATION:  /product= "Module 2 of SorB"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  30881..35446
;      OTHER INFORMATION:  /product= "Module 3 of SorB"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  35528..40114
;      OTHER INFORMATION:  /product= "Module 4 of SorB"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  40190..46318
;      OTHER INFORMATION:  /product= "Module 5 of SorB"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  46851..47891
;      OTHER INFORMATION:  /product= "SorM"
;      OTHER INFORMATION:  /note= "The protein encoded by the sorM gene is highly
;      OTHER INFORMATION:  homologous to the methyltransferase from Streptomyces
;      OTHER INFORMATION:  hygroscopicus that is involved in the synthesis of the
;      OTHER INFORMATION:  polyketide rappamicin."
US-08-764-233A-1

  Query Match             24.1%;  Score 105.6;  DB 1;  Length 49377;
  Best Local Similarity   56.7%;  Pred. No. 3.1e-16;
  Matches  195;  Conservative    0;  Mismatches  149;  Indels    0;  Gaps    0;

Qy          1 TACTGGCTCGAAAGCACACAGCCCGGTGCCGGCAACGTGTCAGCAGCCGGACTCGACCCC 60
              |||||||||   | |  || |     || | | ||||  |  | || || ||   | |
Db       9855 TTCTGGCTCGATGCCTCCAAGGCACACGCTGCCGACGTCGCCTCCGCAGGCCTGACCTCG 9914

Qy         61 ACCGAACACCCCCTACTCGGCGCCACATTGGAACTGGCGACTGACGGTGGAGCGCTTCTT 120
              ||||| ||||| || |||||||||     |    || ||       || |||    ||  |
Db       9915 ACCGATCACCCGCTGCTCGGCGCCGGCGTCCCCCTCGCCGACCGCGATGGCTTCCTCTTC 9974

Qy        121 GCAGGGCGCTTGTCTTTGAGGTCGCATCCGTGGCTGGCTGACCATGCCGTCGGCGGCACG 180
               |||| ||  | ||  |      ||||||||||| || || || | ||||  ||| ||
Db       9975 ACAGGACGACTCTCACTCTCAGAGCATCCGTGGCTCGCCGATCACGTCGTCTTCGGTACA 10034

Qy        181 GTGCTGCTGTCGGGCGCCACCTTCCTCGAACTCGCCCTTCATGCGGGCACATACGTGGGC 240
                | ||  |||||| |  |||| ||||| || |||||    |  | |     ||| |||
Db      10035 CCCATCCTTCCGGGCACTGCCTTTCTCGAGCTTGCCCTGTTCGTCGCCGGTCGCGTCGGC 10094

Qy        241 TGCGACCGAGTGGATGAGCTGACGCTGCATGCGCCGCTGGTGGTTCCTGTGGATGGGGGT 300
               ||||   || || ||||| || ||  |  | || || |   | ||    || || |
Db      10095 CTCGACACCGTCGAAGAGCTCACCCTCGAAACCCCCCTCGCTCTCCCGTCTGAAGGCGCC 10154

Qy        301 GTGAGTGTGCAGGTTGGGGTTGCGGCTGCGGATGGGGAGGGGCG 344
               |    || |||||   ||| | | ||  ||| |  |  || ||
Db      10155 CTCCTCGTCCAGGTGTCGGTCGGGCCTTTGGACGACGCAGGACG 10198


Search completed: June 18, 2003, 01:07:50
Job time : 36.9737 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 17, 2003, 11:26:45 ; Search time 87.13 Seconds
(without alignments)
11320.736 Million cell updates/sec

Title: US-09-914-286-1_COPY_8947_9384
Perfect score: 438
Sequence: 1 tactggctcgaaagcacaca.....gtcatgcctcgggggtgctg 438

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_101002:*
1: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	Query Match	Length	DB	ID	Description
	1	438	100.0	11916	22	AAH79279	Streptomyces averm
	2	438	100.0	12381	21	AAZ58381	Streptomyces averm
	3	438	100.0	30690	21	AAA92301	S. avermitilis ave
	4	438	100.0	30690	22	AAH79277	Streptomyces averm
	5	431.6	98.5	438	21	AAA92303	S. avermitilis ave
	6	273.6	62.5	31422	21	AAA92302	S. avermitilis ave
	7	273.6	62.5	31422	22	AAH79278	Streptomyces averm
	8	133.6	30.5	11220	21	AAZ87298	S. venezuelae macr
	9	133.6	30.5	36778	21	AAZ87318	S. venezuelae pik
	10	133.6	30.5	37948	21	AAZ87285	S. venezuelae pik
	11	133.6	30.5	38506	21	AAA75633	Nucleotide sequenc
	12	133.6	30.5	38506	21	AAZ56001	Recombinant cosmid
	13	132.8	30.3	65140	22	AAD17184	Streptomyces nours
	14	132.8	30.3	125401	22	AAD17186	Streptomyces nours
	15	131.2	30.0	27541	22	AAD17185	Streptomyces nours
	16	130.4	29.8	77536	21	AAA14651	Nucleotide sequenc
	17	129.4	29.5	20394	22	AAF24892	Pimaricin biosynth
c	18	124	28.3	1590	24	AAS18432	Contig 143a DNA en
	19	123.4	28.2	16767	22	AAF88339	S. spinosa DNA fra
	20	123.4	28.2	50000	22	AAF88313	S. spinosa DNA fra
	21	123.4	28.2	50000	22	AAF88316	S. spinosa DNA fra
	22	123.4	28.2	80161	20	AAZ21501	DNA fragment of Sa
	23	122.2	27.9	43280	18	AAT80413	Tylactone synthase
	24	118.4	27.0	29736	22	AAF88317	S. spinosa DNA fra
	25	117.4	26.8	50937	21	AAA09469	Streptococcus olea
c	26	114.4	26.1	77536	21	AAA14651	Nucleotide sequenc
	27	111.8	25.5	13987	18	AAT80415	Hybrid srmG/tylG O
	28	111.8	25.5	44377	18	AAT78508	Platenolide syntha
	29	111.8	25.5	44377	18	AAT80414	Platenolide syntha
	30	109.4	25.0	33529	23	AAS17367	DNA sequence of S.
	31	107	24.4	5676	19	AAV21186	Amycolatopsis medi
	32	107	24.4	53789	19	AAV21187	Amycolatopsis medi
	33	106	24.2	6459	22	AAF88336	S. spinosa DNA fra
	34	105.8	24.2	4851	22	AAH52062	Mycobacterium tube
c	35	105.8	24.2	4403765	22	AAI99683	Mycobacterium tube
c	36	105.8	24.2	4411529	22	AAI99682	Mycobacterium tube
	37	105.6	24.1	49377	19	AAV05287	The soraphen biosy
	38	105.2	24.0	15872	18	AAT68715	Streptomyces venez
	39	105.2	24.0	15872	21	AAZ87283	S. venezuelae vep
	40	104.8	23.9	13842	21	AAZ87297	S. venezuelae macr
	41	101.2	23.1	784	15	AAQ44795	Constituent of the
	42	99	22.6	28598	17	AAT06769	Sorangium cellulos
	43	99	22.6	28958	18	AAT89956	Sorangium cellulos
	44	99	22.6	28958	21	AAA75299	DNA sequence of So
	45	96.2	22.0	14775	22	AAF88338	S. spinosa DNA fra

ALIGNMENTS

RESULT 1
AAH79279
ID AAH79279 standard; DNA; 11916 BP.
XX
AC AAH79279;
XX
DT 04-DEC-2001 (first entry)
XX
DE Streptomyces avermitilis coding sequence derivative SEQ ID NO: 3.
XX
KW Avermectin aglycone synthase; AAS; avermectin derivative;
KW drug production; vetinary drug; pesticide; ds.
XX
OS Synthetic.
XX
FH Key Location/Qualifiers
FT CDS 1..11916
FT /*tag= a
FT /product= "AAG65268"
FT /partial
XX
PN WO200162939-A1.
XX
PD 30-AUG-2001.
XX
PF 23-FEB-2001; 2001WO-JP01381.
XX
PR 24-FEB-2000; 2000JP-0047405.
XX
PA (KYOW) KYOWA HAKKO KOGYO KK.
PA (KITA) KITASATO INST.
XX

PI Endo H, Yamaguchi H, Kanda Y, Hashimoto S, Omura S, Ikeda H;
XX
DR WPI; 2001-582053/65.
DR P-PSDB; AAG65268.
XX
PT New modified avermectin aglycone synthase derived from Streptomyces
PT avermectilis used in production of 22,23-dihydroavermectin B1a used in
PT drugs and pesticides -
XX
PS Claim 10; Page 149-167; 257pp; Japanese.
XX
CC The present invention relates to the production of modified derivatives
CC of avermectin aglycone synthase (AAS) derived from Streptomyces
CC avermitilis. The activity of an acyl carrier protein (ACP),
CC beta-ketoacyl carrier protein synthase (KS), acyltransferase (AT),
CC beta-ketoacyl carrier protein reductase (KR), dehydratase (DH), enoyl
CC reductase (ER) and/or thioesterase (TE) domain may be reduced or
CC suppressed. The process can be used in the production of drugs, vetinary
CC drugs and pesticides. The present sequence is modified version of a
CC fragment of the S. avermitilis genome.
XX
SQ Sequence 11916 BP; 1831 A; 4428 C; 3820 G; 1837 T; 0 other;

Query Match 100.0%; Score 438; DB 22; Length 11916;
Best Local Similarity 100.0%; Pred. No. 3.7e-89;
Matches 438; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 TACTGGCTCGAAAGCACACAGCCCGGTGCCGGCAACGTGTCAGCAGCCGGACTCGACCCC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8947 TACTGGCTCGAAAGCACACAGCCCGGTGCCGGCAACGTGTCAGCAGCCGGACTCGACCCC 9006

Qy         61 ACCGAACACCCCCTACTCGGCGCCACATTGGAACTGGCGACTGACGGTGGAGCGCTTCTT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9007 ACCGAACACCCCCTACTCGGCGCCACATTGGAACTGGCGACTGACGGTGGAGCGCTTCTT 9066

Qy        121 GCAGGGCGCTTGTCTTTGAGGTCGCATCCGTGGCTGGCTGACCATGCCGTCGGCGGCACG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9067 GCAGGGCGCTTGTCTTTGAGGTCGCATCCGTGGCTGGCTGACCATGCCGTCGGCGGCACG 9126

Qy        181 GTGCTGCTGTCGGGCGCCACCTTCCTCGAACTCGCCCTTCATGCGGGCACATACGTGGGC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9127 GTGCTGCTGTCGGGCGCCACCTTCCTCGAACTCGCCCTTCATGCGGGCACATACGTGGGC 9186

Qy        241 TGCGACCGAGTGGATGAGCTGACGCTGCATGCGCCGCTGGTGGTTCCTGTGGATGGGGGT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9187 TGCGACCGAGTGGATGAGCTGACGCTGCATGCGCCGCTGGTGGTTCCTGTGGATGGGGGT 9246

Qy        301 GTGAGTGTGCAGGTTGGGGTTGCGGCTGCGGATGGGGAGGGGCGGCGTTTGGTGAGTGTG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9247 GTGAGTGTGCAGGTTGGGGTTGCGGCTGCGGATGGGGAGGGGCGGCGTTTGGTGAGTGTG 9306

Qy        361 TATGCGCGGGGGTGGGAGTGCTTGTGGTGGGGGTGGTGCGTCGGGTGGGGTGTGGACGTGT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9307 TATGCGCGGGGTGGGAGTGCTTGTGGTGGGGGTGGTGCGTCGGGTGGGGTGTGGACGTGT 9366

Qy        421 CATGCCTCGGGGGTGCTG 438
              ||||||||||||||||||
Db       9367 CATGCCTCGGGGGTGCTG 9384
```

RESULT 2
AAZ58381
ID AAZ58381 standard; DNA; 12381 BP.
XX
AC AAZ58381;
XX
DT 23-MAY-2000 (first entry)
XX
DE Streptomyces avermitilis avermectin polyketide synthase modules 1+2.
XX
KW Polyketide synthase; avermectin; insecticide; ss.
XX
OS Streptomyces avermitilis.
XX
PN WO200001827-A2.
XX
PD 13-JAN-2000.
XX
PF 06-JUL-1999; 99WO-GB02158.
XX
PR 06-JUL-1998; 98GB-0014622.
XX
PA (BIOT-) BIOTICA TECHNOLOGY LTD.
PA (PFIZ) PFIZER INC.
XX
PI Kellenberger JL, Leadlay PF, Staunton J, Stutzman-Engwall KJ;
PI McArthur HAI;
XX
DR WPI; 2000-182117/16.
XX
PT Mutated Type I polyketide synthase containing a polylinker site in an
PT extension module for replacement of a reductive loop sequence, for
PT producing polyketides, e.g. B1 avermectin -
XX
PS Disclosure; Fig 7a-f; 75pp; English.
XX
CC The present sequence is that of DNA encoding the first 2 modules
CC of the avermectin polyketide synthase (PKS) of Streptomyces
CC avermitilis. The invention relates to nucleic acids encoding a
CC Type I PKS such as avermectin in which a polylinker with multiple
CC restriction sites replaces or 1 more PKS genes encoding enzymes
CC associated with reduction. Novel PKS are provided in which in
CC which the reductive loop in a selected module of the Type I PKS is
CC replaced with the equivalent segment from the same or different
CC PKS gene cluster or by a mutated or synthetic segment. Vectors and
CC host cells, and methods for producing novel polyketides by
CC culturing host cells are claimed. The polyketides obtained are
CC useful as antibiotics and insecticides. Fermentation products
CC containing C22-C23 dihydroavermectin, ivermectin and B1
CC avermectins are claimed.
XX
SQ Sequence 12381 BP; 1884 A; 4561 C; 4005 G; 1931 T; 0 other;

Query Match 100.0%; Score 438; DB 21; Length 12381;
Best Local Similarity 100.0%; Pred. No. 3.7e-89;
Matches 438; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 TACTGGCTCGAAAGCACACAGCCCGGTGCCGGCAACGTGTCAGCAGCCGGACTCGACCCC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9764 TACTGGCTCGAAAGCACACAGCCCGGTGCCGGCAACGTGTCAGCAGCCGGACTCGACCCC 9823

Qy         61 ACCGAACACCCCCTACTCGGCGCCACATTGGAACTGGCGACTGACGGTGGAGCGCTTCTT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9824 ACCGAACACCCCCTACTCGGCGCCACATTGGAACTGGCGACTGACGGTGGAGCGCTTCTT 9883

Qy        121 GCAGGGCGCTTGTCTTTGAGGTCGCATCCGTGGCTGGCTGACCATGCCGTCGGCGGCACG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9884 GCAGGGCGCTTGTCTTTGAGGTCGCATCCGTGGCTGGCTGACCATGCCGTCGGCGGCACG 9943

Qy        181 GTGCTGCTGTCGGGCGCCACCTTCCTCGAACTCGCCCTTCATGCGGGCACATACGTGGGC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9944 GTGCTGCTGTCGGGCGCCACCTTCCTCGAACTCGCCCTTCATGCGGGCACATACGTGGGC 10003

Qy        241 TGCGACCGAGTGGATGAGCTGACGCTGCATGCGCCGCTGGTGGTTCCTGTGGATGGGGGT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10004 TGCGACCGAGTGGATGAGCTGACGCTGCATGCGCCGCTGGTGGTTCCTGTGGATGGGGGT 10063

Qy        301 GTGAGTGTGCAGGTTGGGGTTGCGGCTGCGGATGGGGAGGGGCGGCGTTTGGTGAGTGTG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10064 GTGAGTGTGCAGGTTGGGGTTGCGGCTGCGGATGGGGAGGGGCGGCGTTTGGTGAGTGTG 10123

Qy        361 TATGCGCGGGGTGGGAGTGCTTGTGGTGGGGGTGGTGCGTCGGGTGGGGTGTGGACGTGT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10124 TATGCGCGGGGTGGGAGTGCTTGTGGTGGGGGTGGTGCGTCGGGTGGGGTGTGGACGTGT 10183

Qy        421 CATGCCTCGGGGGTGCTG 438
```

```
                  ||||||||||||||||||
Db      10184 CATGCCTCGGGGGTGCTG 10201


RESULT 3
AAA92301
ID   AAA92301 standard; DNA; 30690 BP.
XX
AC   AAA92301;
XX
DT   10-JAN-2001  (first entry)
XX
DE   S. avermitilis avermectin aglycon synthase DNA aveAI SEQ ID NO:1.
XX
KW   Streptomyces avermitilis; avermectin aglycon synthase; biosynthesis;
KW   multifunctional enzyme; polyketide; avermectin; veterinary drug;
KW   agrochemical; ds.
XX
OS   Streptomyces avermitilis.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..11919
FT                   /*tag= a
FT                   /note= "avermectin aglycon synthase protein"
FT   CDS             11971..30690
FT                   /*tag= b
FT                   /note= "avermectin aglycon synthase protein"
XX
PN   WO200050605-A1.
XX
PD   31-AUG-2000.
XX
PF   23-FEB-2000; 2000WO-JP01041.
XX
PR   24-FEB-1999;   99JP-0046961.
XX
PA   (KITA ) KITASATO INST.
XX
PI   Omura S,  Ikeda H;
XX
DR   WPI; 2000-565458/52.
DR   P-PSDB; AAB23749, AAB23750.
XX
PT   Avermectin aglycone synthase DNA and proteins encoded by all or part of
PT   it for the production of avermectin and its derivatives for drug and
PT   agrochemical use -
XX
PS   Claim 2; Page 66-134; 314pp; Japanese.
XX
CC   The present sequence represents DNA which encodes avermectin aglycon
CC   synthase proteins. Also described are: (1) polypeptides encoded by all
CC   or part of the DNA; (2) expression vectors containing the DNA; (3) host
CC   cells transformed by the vectors; (4) preparation of the polypeptides
CC   by culture of the transformants; (5) preparation of avermectin aglycon
CC   or its derivatives by culture of transformed avermectin-producing
CC   microorganisms; and (6) oligonucleotides of 5-60 bases in length
CC   containing sense or antisense sequences from the avermectin aglycon
CC   synthase DNA. The enzymes are useful for the production of modified
CC   forms of avermectin and of the intermediates in its biosynthesis, for
CC   use as drugs, veterinary drugs and agrochemicals.
XX
SQ   Sequence 30690 BP; 5356 A; 12454 C; 8617 G; 4263 T; 0 other;

  Query Match            100.0%;  Score 438;  DB 21;  Length 30690;
  Best Local Similarity  100.0%;  Pred. No. 4e-89;
  Matches  438;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 TACTGGCTCGAAAGCACACAGCCCGGTGCCGGCAACGTGTCAGCAGCCGGACTCGACCCC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8947 TACTGGCTCGAAAGCACACAGCCCGGTGCCGGCAACGTGTCAGCAGCCGGACTCGACCCC 9006

Qy         61 ACCGAACACCCCCTACTCGGCGCCACATTGGAACTGGCGACTGACGGTGGAGCGCTTCTT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9007 ACCGAACACCCCCTACTCGGCGCCACATTGGAACTGGCGACTGACGGTGGAGCGCTTCTT 9066

Qy        121 GCAGGGCGCTTGTCTTTGAGGTCGCATCCGTGGCTGGCTGACCATGCCGTCGGCGGCACG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9067 GCAGGGCGCTTGTCTTTGAGGTCGCATCCGTGGCTGGCTGACCATGCCGTCGGCGGCACG 9126

Qy        181 GTGCTGCTGTCGGGCGCCACCTTCCTCGAACTCGCCCTTCATGCGGGCACATACGTGGGC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9127 GTGCTGCTGTCGGGCGCCACCTTCCTCGAACTCGCCCTTCATGCGGGCACATACGTGGGC 9186

Qy        241 TGCGACCGAGTGGATGAGCTGACGCTGCATGCGCCGCTGGTGGTTCCTGTGGATGGGGGT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9187 TGCGACCGAGTGGATGAGCTGACGCTGCATGCGCCGCTGGTGGTTCCTGTGGATGGGGGT 9246

Qy        301 GTGAGTGTGCAGGTTGGGGTTGCGGCTGCGGATGGGGAGGGGCGGCGTTTGGTGAGTGTG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9247 GTGAGTGTGCAGGTTGGGGTTGCGGCTGCGGATGGGGAGGGGCGGCGTTTGGTGAGTGTG 9306

Qy        361 TATGCGCGGGGGTGGGAGTGCTTGTGGTGGGGGTGGTGCGTCGGGTGGGGTGTGGACGTGT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9307 TATGCGCGGGGGTGGGAGTGCTTGTGGTGGGGGTGGTGCGTCGGGTGGGGTGTGGACGTGT 9366

Qy        421 CATGCCTCGGGGGTGCTG 438
              ||||||||||||||||||
Db       9367 CATGCCTCGGGGGTGCTG 9384


RESULT 4
AAH79277
ID   AAH79277 standard; DNA; 30690 BP.
XX
AC   AAH79277;
XX
DT   04-DEC-2001  (first entry)
XX
DE   Streptomyces avermitilis coding sequences SEQ ID NO: 1.
XX
KW   Avermectin aglycone synthase; AAS; avermectin derivative;
KW   drug production; vetinary drug; pesticide; ds.
XX
OS   Streptomyces avermitilis.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..11919
FT                   /*tag= a
FT                   /product= "AAG65264"
FT                   /partial
FT   CDS             11971..30690
FT                   /*tag= b
FT                   /product= "AAG65265"
XX
PN   WO200162939-A1.
XX
PD   30-AUG-2001.
XX
PF   23-FEB-2001; 2001WO-JP01381.
XX
PR   24-FEB-2000; 2000JP-0047405.
XX
PA   (KYOW ) KYOWA HAKKO KOGYO KK.
PA   (KITA ) KITASATO INST.
XX
PI   Endo H,  Yamaguchi H,  Kanda Y,  Hashimoto S,  Omura S,  Ikeda H;
XX
DR   WPI; 2001-582053/65.
DR   P-PSDB; AAG65264, AAG65265.
XX
PT   New modified avermectin aglycone synthase derived from Streptomyces
PT   avermectilis used in production of 22,23-dihydroavermectin B1a used in
PT   drugs and pesticides -
XX
PS   Example 2; Page 58-123; 257pp; Japanese.
XX
```

```
CC   The present invention relates to the production of modified derivatives
CC   of avermectin aglycone synthase (AAS) derived from Streptomyces
CC   avermitilis. The activity of an acyl carrier protein (ACP),
CC   beta-ketoacyl carrier protein synthase (KS), acyltransferase (AT),
CC   beta-ketoacyl carrier protein reductase (KR), dehydratase (DH), enoyl
CC   reductase (ER) and/or thioesterase (TE) domain may be reduced or
CC   suppressed. The process can be used in the production of drugs, vetinary
CC   drugs and pesticides. The present sequence is a fragment of the S.
CC   avermitilis genome.
XX
SQ   Sequence 30690 BP; 5355 A; 12455 C; 8617 G; 4263 T; 0 other;

  Query Match           100.0%;  Score 438;  DB 22;  Length 30690;
  Best Local Similarity  100.0%;  Pred. No. 4e-89;
  Matches  438;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 TACTGGCTCGAAAGCACACAGCCCGGTGCCGGCAACGTGTCAGCAGCCGGACTCGACCCC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8947 TACTGGCTCGAAAGCACACAGCCCGGTGCCGGCAACGTGTCAGCAGCCGGACTCGACCCC 9006

Qy         61 ACCGAACACCCCCTACTCGGCGCCACATTGGAACTGGCGACTGACGGTGGAGCGCTTCTT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9007 ACCGAACACCCCCTACTCGGCGCCACATTGGAACTGGCGACTGACGGTGGAGCGCTTCTT 9066

Qy        121 GCAGGGCGCTTGTCTTTGAGGTCGCATCCGTGGCTGGCTGACCATGCCGTCGGCGGCACG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9067 GCAGGGCGCTTGTCTTTGAGGTCGCATCCGTGGCTGGCTGACCATGCCGTCGGCGGCACG 9126

Qy        181 GTGCTGCTGTCGGGCGCCACCTTCCTCGAACTCGCCCTTCATGCGGGCACATACGTGGGC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9127 GTGCTGCTGTCGGGCGCCACCTTCCTCGAACTCGCCCTTCATGCGGGCACATACGTGGGC 9186

Qy        241 TGCGACCGAGTGGATGAGCTGACGCTGCATGCGCCGCTGGTGGTTCCTGTGGATGGGGGT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9187 TGCGACCGAGTGGATGAGCTGACGCTGCATGCGCCGCTGGTGGTTCCTGTGGATGGGGGT 9246

Qy        301 GTGAGTGTGCAGGTTGGGGTTGCGGCTGCGGATGGGGAGGGGCGGCGTTTGGTGAGTGTG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9247 GTGAGTGTGCAGGTTGGGGTTGCGGCTGCGGATGGGGAGGGGCGGCGTTTGGTGAGTGTG 9306

Qy        361 TATGCGCGGGGTGGGAGTGCTTGTGGTGGGGGTGGTGCGTCGGGTGGGGTGTGGACGTGT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9307 TATGCGCGGGGTGGGAGTGCTTGTGGTGGGGGTGGTGCGTCGGGTGGGGTGTGGACGTGT 9366

Qy        421 CATGCCTCGGGGGTGCTG 438
              ||||||||||||||||||
Db       9367 CATGCCTCGGGGGTGCTG 9384


RESULT 5
AAA92303
ID   AAA92303 standard; DNA; 438 BP.
XX
AC   AAA92303;
XX
DT   10-JAN-2001  (first entry)
XX
DE   S. avermitilis avermectin aglycon synthase DNA SEQ ID NO:7.
XX
KW   Streptomyces avermitilis; avermectin aglycon synthase; biosynthesis;
KW   multifunctional enzyme; polyketide; avermectin; veterinary drug;
KW   agrochemical; ds.
XX
OS   Streptomyces avermitilis.
XX
PN   WO200050605-A1.
XX
PD   31-AUG-2000.
XX
PF   23-FEB-2000; 2000WO-JP01041.
XX
PR   24-FEB-1999;   99JP-0046961.
XX
PA   (KITA ) KITASATO INST.
XX
PI   Omura S,  Ikeda H;
XX
DR   WPI; 2000-565458/52.
DR   P-PSDB; AAB23753.
XX
PT   Avermectin aglycone synthase DNA and proteins encoded by all or part of
PT   it for the production of avermectin and its derivatives for drug and
PT   agrochemical use -
XX
PS   Claim 17; Page 306-307; 314pp; Japanese.
XX
CC   The present sequence represents DNA which encodes an avermectin aglycon
CC   synthase protein. Also described are: (1) polypeptides encoded by all
CC   or part of the DNA; (2) expression vectors containing the DNA; (3) host
CC   cells transformed by the vectors; (4) preparation of the polypeptides
CC   by culture of the transformants; (5) preparation of avermectin aglycon
CC   or its derivatives by culture of transformed avermectin-producing
CC   microorganisms; and (6) oligonucleotides of 5-60 bases in length
CC   containing sense or antisense sequences from the avermectin aglycon
CC   synthase DNA. The enzymes are useful for the production of modified
CC   forms of avermectin and of the intermediates in its biosynthesis, for
CC   use as drugs, veterinary drugs and agrochemicals.
XX
SQ   Sequence 438 BP; 55 A; 112 C; 178 G; 93 T; 0 other;

  Query Match           98.5%;  Score 431.6;  DB 21;  Length 438;
  Best Local Similarity  99.1%;  Pred. No. 7.3e-88;
  Matches  434;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy          1 TACTGGCTCGAAAGCACACAGCCCGGTGCCGGCAACGTGTCAGCAGCCGGACTCGACCCC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 TACTGGCTCGAAAGCACACAGCCCGGTGCCGGCAACGTGTCAGCAGCCGGACTCGACCCC 60

Qy         61 ACCGAACACCCCCTACTCGGCGCCACATTGGAACTGGCGACTGACGGTGGAGCGCTTCTT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ACCGAACACCCCCTACTCGGCGCCACATTGGAACTGGCGACTGACGGTGGAGCGCTTCTT 120

Qy        121 GCAGGGCGCTTGTCTTTGAGGTCGCATCCGTGGCTGGCTGACCATGCCGTCGGCGGCACG 180
              |||||||||||||||||||||||||||||||||||||||||| | |  ||||||||||||
Db        121 GCAGGGCGCTTGTCTTTGAGGTCGCATCCGTGGCTGGCTGACTACGAGGTCGGCGGCACG 180

Qy        181 GTGCTGCTGTCGGGCGCCACCTTCCTCGAACTCGCCCTTCATGCGGGCACATACGTGGGC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GTGCTGCTGTCGGGCGCCACCTTCCTCGAACTCGCCCTTCATGCGGGCACATACGTGGGC 240

Qy        241 TGCGACCGAGTGGATGAGCTGACGCTGCATGCGCCGCTGGTGGTTCCTGTGGATGGGGGT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TGCGACCGAGTGGATGAGCTGACGCTGCATGCGCCGCTGGTGGTTCCTGTGGATGGGGGT 300

Qy        301 GTGAGTGTGCAGGTTGGGGTTGCGGCTGCGGATGGGGAGGGGCGGCGTTTGGTGAGTGTG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GTGAGTGTGCAGGTTGGGGTTGCGGCTGCGGATGGGGAGGGGCGGCGTTTGGTGAGTGTG 360

Qy        361 TATGCGCGGGGTGGGAGTGCTTGTGGTGGGGGTGGTGCGTCGGGTGGGGTGTGGACGTGT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 TATGCGCGGGGTGGGAGTGCTTGTGGTGGGGGTGGTGCGTCGGGTGGGGTGTGGACGTGT 420

Qy        421 CATGCCTCGGGGGTGCTG 438
              ||||||||||||||||||
Db        421 CATGCCTCGGGGGTGCTG 438


RESULT 6
AAA92302
ID   AAA92302 standard; DNA; 31422 BP.
XX
AC   AAA92302;
XX
DT   10-JAN-2001  (first entry)
XX
DE   S. avermitilis avermectin aglycon synthase DNA aveAII SEQ ID NO:2.
```

```
XX
KW   Streptomyces avermitilis; avermectin aglycon synthase; biosynthesis;
KW   multifunctional enzyme; polyketide; avermectin; veterinary drug;
KW   agrochemical; ds.
XX
OS   Streptomyces avermitilis.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..14646
FT                   /*tag= a
FT                   /note= "avermectin aglycon synthase protein"
FT   CDS             14824..31422
FT                   /*tag= b
FT                   /note= "avermectin aglycon synthase protein"
XX
PN   WO200050605-A1.
XX
PD   31-AUG-2000.
XX
PF   23-FEB-2000; 2000WO-JP01041.
XX
PR   24-FEB-1999;   99JP-0046961.
XX
PA   (KITA ) KITASATO INST.
XX
PI   Omura S,  Ikeda H;
XX
DR   WPI; 2000-565458/52.
DR   P-PSDB; AAB23751, AAB23752.
XX
PT   Avermectin aglycone synthase DNA and proteins encoded by all or part of
PT   it for the production of avermectin and its derivatives for drug and
PT   agrochemical use -
XX
PS   Claim 2; Page 134-203; 314pp; Japanese.
XX
CC   The present sequence represents DNA which encodes avermectin aglycon
CC   synthase proteins. Also described are: (1) polypeptides encoded by all
CC   or part of the DNA; (2) expression vectors containing the DNA; (3) host
CC   cells transformed by the vectors; (4) preparation of the polypeptides
CC   by culture of the transformants; (5) preparation of avermectin aglycon
CC   or its derivatives by culture of transformed avermectin-producing
CC   microorganisms; and (6) oligonucleotides of 5-60 bases in length
CC   containing sense or antisense sequences from the avermectin aglycon
CC   synthase DNA. The enzymes are useful for the production of modified
CC   forms of avermectin and of the intermediates in its biosynthesis, for
CC   use as drugs, veterinary drugs and agrochemicals.
XX
SQ   Sequence 31422 BP; 4136 A; 10238 C; 11677 G; 5371 T; 0 other;

  Query Match             62.5%;  Score 273.6;  DB 21;  Length 31422;
  Best Local Similarity   78.6%;  Pred. No. 2.9e-52;
  Matches  327;  Conservative    0;  Mismatches   89;  Indels    0;  Gaps    0;

Qy         23 CCGGTGCCGGCAACGTGTCAGCAGCCGGACTCGACCCCACCGAACACCCCCTACTCGGCG 82
              | || || ||  ||||  | || || ||  | || ||  |||||||||| || ||||  |
Db      17705 CGGGAGCAGGTGACGTCGCCGCTGCTGGCTTGGAGCCGGCCGAACACCCTCTGCTCGCGG 17764

Qy         83 CCACATTGGAACTGGCGACTGACGGTGGAGCGCTTCTTGCAGGGCGCTTGTCTTTGAGGT 142
              | ||| |  |||| ||        |  ||    || ||  | ||  ||| |||| ||| | |
Db      17765 CAACAGTCCAACTCGCAGACACGGACGGCTGCCTACTGACGGGTCGCCTGTCCTTGCGCT 17824

Qy        143 CGCATCCGTGGCTGGCTGACCATGCCGTCGGCGGCACGGTGCTGCTGTCGGGCGCCACCT 202
              ||||||||||||||  ||  | |  || || ||  |||| ||||||||||  |  | |
Db      17825 CGCATCCGTGGCTGGGCGATTACGAGGTGGGGGGTGCGGTCCTGCTGTCGGGGTCGGCGT 17884

Qy        203 TCCTCGAACTCGCCCTTCATGCGGGCACATACGTGGGCTGCGACCGAGTGGATGAGCTGA 262
              || | || || ||  | || |  |||  |  ||||||||||  |||| | ||  | || |
Db      17885 TCGTGGAGCTGGCGGTCCAGGTTGGCGAACGCGTGGGCTGCACCCGAATCGAGCAACTCA 17944

Qy        263 CGCTGCATGCGCCGCTGGTGGTTCCTGTGGATGGGGGTGTGAGTGTGCAGGTTGGGGTTG 322
              | |||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db      17945 CTGTGCATGCGCCGCTGGTGGTTCCTGTGGGTGGGGGTGTGAGTGTGCAGGTTGGGGTTG 18004

Qy        323 CGGCTGCGGATGGGGAGGGGCGGCGTTTGGTGAGTGTGTATGCGCGGGGTGGGAGTGCTT 382
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      18005 CGGCTGCGGATGGGGAGGGGCGGCGTTTGGTGAGTGTGTATGCGCGGGGTGGGAGTGCTT 18064

Qy        383 GTGGTGGGGGTGGTGCGTCGGGTGGGGTGTGGACGTGTCATGCCTCGGGGGTGCTG 438
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      18065 GTGGTGGGGGTGGTGCGTCGGGTGGGGTGTGGACGTGTCATGCCTCGGGGGTGCTG 18120


RESULT 7
AAH79278
ID   AAH79278 standard; DNA; 31422 BP.
XX
AC   AAH79278;
XX
DT   04-DEC-2001  (first entry)
XX
DE   Streptomyces avermitilis coding sequences SEQ ID NO: 2.
XX
KW   Avermectin aglycone synthase; AAS; avermectin derivative;
KW   drug production; vetinary drug; pesticide; ds.
XX
OS   Streptomyces avermitilis.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..14646
FT                   /*tag= a
FT                   /product= "AAG65266"
FT   CDS             14824..31422
FT                   /*tag= b
FT                   /product= "AAG65267"
XX
PN   WO200162939-A1.
XX
PD   30-AUG-2001.
XX
PF   23-FEB-2001; 2001WO-JP01381.
XX
PR   24-FEB-2000; 2000JP-0047405.
XX
PA   (KYOW ) KYOWA HAKKO KOGYO KK.
PA   (KITA ) KITASATO INST.
XX
PI   Endo H,  Yamaguchi H,  Kanda Y,  Hashimoto S,  Omura S,  Ikeda H;
XX
DR   WPI; 2001-582053/65.
DR   P-PSDB; AAG65266, AAG65267.
XX
PT   New modified avermectin aglycone synthase derived from Streptomyces
PT   avermectilis used in production of 22,23-dihydroavermectin B1a used in
PT   drugs and pesticides -
XX
PS   Disclosure; Page 103-149; 257pp; Japanese.
XX
CC   The present invention relates to the production of modified derivatives
CC   of avermectin aglycone synthase (AAS) derived from Streptomyces
CC   avermitilis. The activity of an acyl carrier protein (ACP),
CC   beta-ketoacyl carrier protein synthase (KS), acyltransferase (AT),
CC   beta-ketoacyl carrier protein reductase (KR), dehydratase (DH), enoyl
CC   reductase (ER) and/or thioesterase (TE) domain may be reduced or
CC   suppressed. The process can be used in the production of drugs, vetinary
CC   drugs and pesticides. The present sequence is a fragment of the S.
CC   avermitilis genome.
XX
SQ   Sequence 31422 BP; 4136 A; 10237 C; 11677 G; 5372 T; 0 other;

  Query Match             62.5%;  Score 273.6;  DB 22;  Length 31422;
  Best Local Similarity   78.6%;  Pred. No. 2.9e-52;
  Matches  327;  Conservative    0;  Mismatches   89;  Indels    0;  Gaps    0;

Qy         23 CCGGTGCCGGCAACGTGTCAGCAGCCGGACTCGACCCCACCGAACACCCCCTACTCGGCG 82
              | || || ||  ||||  | || || ||  | || ||  |||||||||| || ||||  |
```

```
Db      17705 CGGGAGCAGGTGACGTCGCCGCTGCTGGCTTGGAGCCGGCCGAACACCCTCTGCTCGCGG 17764

Qy         83 CCACATTGGAACTGGCGACTGACGGTGGAGCGCTTCTTGCAGGGCGCTTGTCTTTGAGGT 142

Db      17765 CAACAGTCCAACTCGCAGACACGGACGGCTGCCTACTGACGGGTCGCCTGTCCTTGCGCT 17824

Qy        143 CGCATCCGTGGCTGGCTGACCATGCCGTCGGCGGCACGGTGCTGCTGTCGGGCGCCACCT 202

Db      17825 CGCATCCGTGGCTGGGCGATTACGAGGTGGGGGGTGCGGTCCTGCTGTCGGGGTCGGCGT 17884

Qy        203 TCCTCGAACTCGCCCTTCATGCGGGCACATACGTGGGCTGCGACCGAGTGGATGAGCTGA 262

Db      17885 TCGTGGAGCTGGCGGTCCAGGTTGGCGAACGCGTGGGCTGCACCCGAATCGAGCAACTCA 17944

Qy        263 CGCTGCATGCGCCGCTGGTGGTTCCTGTGGATGGGGGTGTGAGTGTGCAGGTTGGGGTTG 322

Db      17945 CTGTGCATGCGCCGCTGGTGGTTCCTGTGGGTGGGGGTGTGAGTGTGCAGGTTGGGGTTG 18004

Qy        323 CGGCTGCGGATGGGGAGGGGCGGCGTTTGGTGAGTGTGTATGCGCGGGGTGGGAGTGCTT 382

Db      18005 CGGCTGCGGATGGGGAGGGGCGGCGTTTGGTGAGTGTGTATGCGCGGGGTGGGAGTGCTT 18064

Qy        383 GTGGTGGGGGTGGTGCGTCGGGTGGGGTGTGGACGTGTCATGCCTCGGGGGTGCTG 438

Db      18065 GTGGTGGGGGTGGTGCGTCGGGTGGGGTGTGGACGTGTCATGCCTCGGGGGTGCTG 18120


RESULT 8
AAZ87298
ID   AAZ87298 standard; DNA; 11220 BP.
XX
AC   AAZ87298;
XX
DT   05-JUN-2000  (first entry)
XX
DE   S. venezuelae macrolide biosynthetic gene pikAII, SEQ ID NO:32.
XX
KW   Desosamine biosynthesis; macrolide; polyketide; methymycin; pikromycin;
KW   neomethymycin; narbomycin; polyhydroxyalkanoate monomer synthase;
KW   biopolymer; antibiotic; chemotherapeutic; immunosuppressant; asthma;
KW   chronic obstructive pulmonary disease; respiratory inflammation;
KW   hypercholesterolaemia; crop protection agent; ds.
XX
OS   Streptomyces venezuelae ATCC15439.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..111220
FT                   /*tag=  a
FT                   /product= "PikAII"
XX
PN   WO200000620-A2.
XX
PD   06-JAN-2000.
XX
PF   25-JUN-1999;   99WO-US14398.
XX
PR   26-JUN-1998;   98US-0105537.
XX
PA   (MINU ) UNIV MINNESOTA.
XX
PI   Sherman DH,  Liu H,  Xue Y,  Zhao L;
XX
DR   WPI; 2000-160679/14.
DR   P-PSDB; AAY77193.
XX
PT   Desosamine and macrolide biosynthetic gene clusters, useful for, e .g.
PT   synthesis of methymycin and pikromycin -
XX
PS   Claim 15; Page 398-403; 438pp; English.
XX
CC   The invention relates to an isolated and purified nucleic acid segment
CC   comprising a desosamine biosynthetic gene cluster, a fragment or its
CC   biologically active variant, where the nucleic acid sequence is not
CC   derived from the eryC gene cluster of Saccharopolyspora erythraea or
CC   Streptomyces antibioticus. The invention also relates to a macrolide
CC   biosynthetic gene cluster, or fragments thereof. The macrolide
CC   biosynthetic gene cluster encodes proteins which synthesise methymycin,
CC   pikromycin, neomethymycin, narbomycin or a combination of these
CC   compounds. Recombinant or augmented cells comprising the desosamine
CC   and/or macrolide biosynthetic gene clusters are useful for the
CC   production of biologically active macrolides. The macrolide biosynthetic
CC   proteins are useful for synthesis of methymycin, pikromycin,
CC   neomethymycin and narbomycin. The alternative termination of polyketide
CC   synthesis may be useful to prepare novel antibiotics and
CC   polyhydroxyalkanoate (PHA) monomers. The compounds produced by the
CC   recombinant host cells are useful as biopolymers, e.g., in packaging or
CC   biomedical applications, to engineer PHA monomer synthases or to prepare
CC   biologically active agents, such as chemotherapeutics,
CC   immunosuppressants, agents to treat asthma, chronic obstructive pulmonary
CC   disease as well as other diseases involving respiratory inflammation,
CC   cholesterol-lowering agents or macrolide-based antibiotics which are
CC   active against a variety of organisms, e.g., bacteria, including
CC   multi-drug resistant pneumococci and other respiratory pathogens, as well
CC   as viral parasitic pathogens, or as crop protection agents (e.g.,
CC   fungicides or insecticides) via expression of polyketides in plants.
CC   Sequences AAZ87295-Z87302 represent macrolide biosynthetic genes from
CC   Streptomyces venezuelae ATCC 15439, which encode proteins
CC   AAY77190-Y77197.
XX
SQ   Sequence 11220 BP; 1369 A; 4423 C; 3966 G; 1462 T; 0 other;

  Query Match           30.5%;  Score 133.6;  DB 21;  Length 11220;
  Best Local Similarity  58.2%;  Pred. No. 6.3e-21;
  Matches  235;  Conservative    0;  Mismatches  169;  Indels    0;  Gaps    0;

Qy          1 TACTGGCTCGAAAGCACACAGCCCGGTGCCGGCAACGTGTCAGCAGCCGGACTCGACCCC 60

Db       7216 TACTGGCCGCAGCCCGACCTCTCCGCCGCCGGTGACATCACCTCCGCCGGTCTCGGGGCG 7275

Qy         61 ACCGAACACCCCCTACTCGGCGCCACATTGGAACTGGCGACTGACGGTGGAGCGCTTCTT 120

Db       7276 GCCGAGCACCCGCTGCTCGGCGCGGCCGTGGCGCTCGCGGACTCCGACGGCTGCCTGCTC 7335

Qy        121 GCAGGGCGCTTGTCTTTGAGGTCGCATCCGTGGCTGGCTGACCATGCCGTCGGCGGCACG 180

Db       7336 ACGGGGAGCCTCTCCCTCCGTACGCACCCCTGGCTGGCGGACCACGCGGTGGCCGGCACC 7395

Qy        181 GTGCTGCTGTCGGGCGCCACCTTCCTCGAACTCGCCCTTCATGCGGGCACATACGTGGGC 240

Db       7396 GTGCTGCTGCCGGGAACGGCGTTCGTGGAGCTGGCGTTCCGAGCCGGGGACCAGGTCGGT 7455

Qy        241 TGCGACCGAGTGGATGAGCTGACGCTGCATGCGCCGCTGGTGGTTCCTGTGGATGGGGGT 300

Db       7456 TGCGATCTGGTCGAGGAGCTCACCCTCGACGCGCCGCTCGTGCTGCCCCGTCGTGGCGCG 7515

Qy        301 GTGAGTGTGCAGGTTGGGGTTGCGGCTGCGGATGGGGAGGGGCGGCGTTTGGTGAGTGTG 360

Db       7516 GTCCGTGTGCAGCTGTCCGTCGGCGCGAGCGACGAGTCCGGGCGTCGTACCTTCGGGCTC 7575

Qy        361 TATGCGCGGGGTGGGAGTGCTTGTGGTGGGGGTGGTGCGTCGGG 404

Db       7576 TACGCGCACCCGGAGGACGCGCCGGGCGAGGCGGAGTGGACGCG 7619


RESULT 9
AAZ87318
ID   AAZ87318 standard; DNA; 36778 BP.
XX
AC   AAZ87318;
XX
DT   05-JUN-2000  (first entry)
XX
DE   S. venezuelae pik (macrolide biosynthesis) gene cluster.
XX
KW   Desosamine biosynthesis; macrolide; polyketide; methymycin; pikromycin;
KW   neomethymycin; narbomycin; polyhydroxyalkanoate monomer synthase;
```

```
KW   biopolymer; antibiotic; chemotherapeutic; immunosuppressant; asthma;
KW   chronic obstructive pulmonary disease; respiratory inflammation;
KW   hypercholesterolaemia; crop protection agent; ds.
XX
OS   Streptomyces venezuelae ATCC15439.
XX
FH   Key             Location/Qualifiers
FT   CDS             1742..15583
FT                   /*tag= a
FT                   /product= "Pik gene cluster protein #1 (AAY77200)"
FT   CDS             15688..26907
FT                   /*tag= b
FT                   /product= "Pik gene cluster protein #2 (AAY77201)"
FT   CDS             26991..31679
FT                   /*tag= c
FT                   /product= "Pik gene cluster protein #3 (AAY77202)"
FT   CDS             31782..35822
FT                   /*tag= d
FT                   /product= "Pik gene cluster protein #4 (AAY77203)"
FT   CDS             35819..36664
FT                   /*tag= e
FT                   /product= "Pik gene cluster protein #4 (AAY80997)"
XX
PN   WO200000620-A2.
XX
PD   06-JAN-2000.
XX
PF   25-JUN-1999;   99WO-US14398.
XX
PR   26-JUN-1998;   98US-0105537.
XX
PA   (MINU ) UNIV MINNESOTA.
XX
PI   Sherman DH,  Liu H,  Xue Y,  Zhao L;
XX
DR   WPI; 2000-160679/14.
DR   P-PSDB; AAY77200, AAY77201, AAY77202, AAY77203, AAY80997.
XX
PT   Desosamine and macrolide biosynthetic gene clusters, useful for, e .g.
PT   synthesis of methymycin and pikromycin -
XX
PS   Disclosure; Figure 31; 438pp; English.
XX
CC   The invention relates to an isolated and purified nucleic acid segment
CC   comprising a desosamine biosynthetic gene cluster, a fragment or its
CC   biologically active variant, where the nucleic acid sequence is not
CC   derived from the eryC gene cluster of Saccharopolyspora erythraea or
CC   Streptomyces antibioticus. The invention also relates to a macrolide
CC   biosynthetic gene cluster, or fragments thereof. The macrolide
CC   biosynthetic gene cluster encodes proteins which synthesise methymycin,
CC   pikromycin, neomethymycin, narbomycin or a combination of these
CC   compounds. Recombinant or augmented cells comprising the desosamine
CC   and/or macrolide biosynthetic gene clusters are useful for the production
CC   of biologically active macrolides. The macrolide biosynthetic proteins
CC   are useful for synthesis of methymycin, pikromycin, neomethymycin and
CC   narbomycin. The alternative termination of polyketide synthesis may be
CC   useful to prepare novel antibiotics and polyhydroxyalkanoate (PHA)
CC   monomers. The compounds produced by the recombinant host cells are useful
CC   as biopolymers, e.g., in packaging or biomedical applications, to
CC   engineer PHA monomer synthases or to prepare biologically active agents,
CC   such as chemotherapeutics, immunosuppressants, agents to treat asthma,
CC   chronic obstructive pulmonary disease as well as other diseases involving
CC   respiratory inflammation, cholesterol-lowering agents or macrolide-based
CC   antibiotics which are active against a variety of organisms, e.g.,
CC   bacteria, including multi-drug resistant pneumococci and other
CC   respiratory pathogens, as well as viral parasitic pathogens, or as crop
CC   protection agents (e.g., fungicides or insecticides) via expression of
CC   polyketides in plants. The present sequence represents the macrolide
CC   biosynthetic gene cluster (pik) from Streptomyces venezuelae ATCC
CC   15439, as given in figure 31.
XX
SQ   Sequence 36778 BP; 4758 A; 14303 C; 12864 G; 4851 T; 2 other;
```

```
  Query Match           30.5%;  Score 133.6;  DB 21;  Length 36778;
  Best Local Similarity  58.2%;  Pred. No. 7.1e-21;
  Matches  235;  Conservative    0;  Mismatches  169;  Indels    0;  Gaps    0;

Qy          1 TACTGGCTCGAAAGCACACAGCCCGGTGCCGGCAACGTGTCAGCAGCCGGACTCGACCCC 60
              ||||||| |  |  |  ||| ||||| || |  |  | ||||| |||| |
Db      22903 TACTGGCCGCAGCCCGACCTCTCCGCCGCCGGTGACATCACCTCCGCCGGTCTCGGGGCG 22962

Qy         61 ACCGAACACCCCCTACTCGGCGCCACATTGGAACTGGCGACTGACGGTGGAGCGCTTCTT 120
              |||| ||||| || |||||||  |  |||  || |||     ||  ||    || ||
Db      22963 GCCGAGCACCCGCTGCTCGGCGCGGCCGTGGCGCTCGCGGACTCCGACGGCTGCCTGCTC 23022

Qy        121 GCAGGGCGCTTGTCTTTGAGGTCGCATCCGTGGCTGGCTGACCATGCCGTCGGCGGCACG 180
              | ||| || | ||  |  |  |||| || |||||||| ||||| || || | ||||||
Db      23023 ACGGGGAGCCTCTCCCTCCGTACGCACCCCTGGCTGGCGGACCACGCGGTGGCCGGCACC 23082

Qy        181 GTGCTGCTGTCGGGCGCCACCTTCCTCGAACTCGCCCTTCATGCGGGCACATACGTGGGC 240
              |||||||| |||  |  | ||| | || || ||  | |  || ||     | || ||
Db      23083 GTGCTGCTGCCGGGAACGGCGTTCGTGGAGCTGGCGTTCCGAGCCGGGGACCAGGTCGGT 23142

Qy        241 TGCGACCGAGTGGATGAGCTGACGCTGCATGCGCCGCTGGTGGTTCCTGTGGATGGGGGT 300
              ||||| |  || || ||||| || ||  | |||||||| ||| | ||     ||| |
Db      23143 TGCGATCTGGTCGAGGAGCTCACCCTCGACGCGCCGCTCGTGCTGCCCCGTCGTGGCGCG 23202

Qy        301 GTGAGTGTGCAGGTTGGGGTTGCGGCTGCGGATGGGGAGGGGCGGCGTTTGGTGAGTGTG 360
              ||  |||||||| |    || |  ||    || | |  ||||| |||    | | |
Db      23203 GTCCGTGTGCAGCTGTCCGTCGGCGCGAGCGACGAGTCCGGGCGTCGTACCTTCGGGCTC 23262

Qy        361 TATGCGCGGGGTGGGAGTGCTTGTGGTGGGGGTGGTGCGTCGGG 404
              || ||||     | |   ||   || | ||  |    | || |
Db      23263 TACGCGCACCCGGAGGACGCGCCGGGCGAGGCGGAGTGGACGCG 23306
```

```
RESULT 10
AAZ87285
ID   AAZ87285 standard; DNA; 37948 BP.
XX
AC   AAZ87285;
XX
DT   05-JUN-2000  (first entry)
XX
DE   S. venezuelae pik (macrolide biosynthesis) gene cluster, SEQ ID NO:5.
XX
KW   Desosamine biosynthesis; macrolide; polyketide; methymycin; pikromycin;
KW   neomethymycin; narbomycin; polyhydroxyalkanoate monomer synthase;
KW   biopolymer; antibiotic; chemotherapeutic; immunosuppressant; asthma;
KW   chronic obstructive pulmonary disease; respiratory inflammation;
KW   hypercholesterolaemia; crop protection agent; ds.
XX
OS   Streptomyces venezuelae ATCC15439.
XX
PN   WO200000620-A2.
XX
PD   06-JAN-2000.
XX
PF   25-JUN-1999;   99WO-US14398.
XX
PR   26-JUN-1998;   98US-0105537.
XX
PA   (MINU ) UNIV MINNESOTA.
XX
PI   Sherman DH,  Liu H,  Xue Y,  Zhao L;
XX
DR   WPI; 2000-160679/14.
DR   P-PSDB; AAY77180.
XX
PT   Desosamine and macrolide biosynthetic gene clusters, useful for, e .g.
PT   synthesis of methymycin and pikromycin -
XX
PS   Claim 13; Page 299-315; 438pp; English.
XX
CC   The invention relates to an isolated and purified nucleic acid segment
CC   comprising a desosamine biosynthetic gene cluster, a fragment or its
```

CC biologically active variant, where the nucleic acid sequence is not
CC derived from the eryC gene cluster of Saccharopolyspora erythraea or
CC Streptomyces antibioticus. The invention also relates to a macrolide
CC biosynthetic gene cluster, or fragments thereof. The macrolide
CC biosynthetic gene cluster encodes proteins which synthesise methymycin,
CC pikromycin, neomethymycin, narbomycin or a combination of these
CC compounds. Recombinant or augmented cells comprising the desosamine
CC and/or macrolide biosynthetic gene clusters are useful for the production
CC of biologically active macrolides. The macrolide biosynthetic proteins
CC are useful for synthesis of methymycin, pikromycin, neomethymycin and
CC narbomycin. The alternative termination of polyketide synthesis may be
CC useful to prepare novel antibiotics and polyhydroxyalkanoate (PHA)
CC monomers. The compounds produced by the recombinant host cells are useful
CC as biopolymers, e.g., in packaging or biomedical applications, to
CC engineer PHA monomer synthases or to prepare biologically active agents,
CC such as chemotherapeutics, immunosuppressants, agents to treat asthma,
CC chronic obstructive pulmonary disease as well as other diseases involving
CC respiratory inflammation, cholesterol-lowering agents or macrolide-based
CC antibiotics which are active against a variety of organisms, e.g.,
CC bacteria, including multi-drug resistant pneumococci and other
CC respiratory pathogens, as well as viral parasitic pathogens, or as crop
CC protection agents (e.g., fungicides or insecticides) via expression of
CC polyketides in plants. The present sequence represents the macrolide
CC biosynthetic gene cluster (pik) from Streptomyces venezuelae ATCC 15439.
XX
SQ Sequence 37948 BP; 4901 A; 14718 C; 13311 G; 5018 T; 0 other;

Query Match 30.5%; Score 133.6; DB 21; Length 37948;
Best Local Similarity 58.2%; Pred. No. 7.1e-21;
Matches 235; Conservative 0; Mismatches 169; Indels 0; Gaps 0;

```
Qy          1 TACTGGCTCGAAAGCACACAGCCCGGTGCCGGCAACGTGTCAGCAGCCGGACTCGACCCC 60
              |||||||   |   |   |   |||  |||||  || |  |  | ||||| ||||   |
Db      24073 TACTGGCCGCAGCCCGACCTCTCCGCCGCCGGTGACATCACCTCCGCCGGTCTCGGGGCG 24132

Qy         61 ACCGAACACCCCCTACTCGGCGCCACATTGGAACTGGCGACTGACGGTGGAGCGCTTCTT 120
               ||||  |||||  || ||||||||  |  |||  ||  |||     ||  ||    || ||
Db      24133 GCCGAGCACCCGCTGCTCGGCGCGGCCGTGGCGCTCGCGGACTCCGACGGCTGCCTGCTC 24192

Qy        121 GCAGGGCGCTTGTCTTTGAGGTCGCATCCGTGGCTGGCTGACCATGCCGTCGGCGGCACG 180
              | ||| || | ||  |  |  |||| || |||||||| ||||| || || | ||||||
Db      24193 ACGGGGAGCCTCTCCCTCCGTACGCACCCCTGGCTGGCGGACCACGCGGTGGCCGGCACC 24252

Qy        181 GTGCTGCTGTCGGGCGCCACCTTCCTCGAACTCGCCCTTCATGCGGGCACATACGTGGGC 240
              |||||||| ||||  |  | ||| | || || ||  | |  || ||      | || ||
Db      24253 GTGCTGCTGCCGGGAACGGCGTTCGTGGAGCTGGCGTTCCGAGCCGGGGACCAGGTCGGT 24312

Qy        241 TGCGACCGAGTGGATGAGCTGACGCTGCATGCGCCGCTGGTGGTTCCTGTGGATGGGGGT 300
              |||||  |  || || ||||| || ||  | |||||||| ||| | ||      ||| |
Db      24313 TGCGATCTGGTCGAGGAGCTCACCCTCGACGCGCCGCTCGTGCTGCCCCGTCGTGGCGCG 24372

Qy        301 GTGAGTGTGCAGGTTGGGGTTGCGGCTGCGGATGGGGAGGGGCGGCGTTTGGTGAGTGTG 360
              ||  ||||||||  |     || |  ||    || | |   ||||| |||     |  | |
Db      24373 GTCCGTGTGCAGCTGTCCGTCGGCGCGAGCGACGAGTCCGGGCGTCGTACCTTCGGGCTC 24432

Qy        361 TATGCGCGGGGTGGGAGTGCTTGTGGTGGGGGTGGTGCGTCGGG 404
              || ||||     | |   ||    || | ||  |    | || |
Db      24433 TACGCGCACCCGGAGGACGCGCCGGGCGAGGCGGAGTGGACGCG 24476
```

RESULT 11
AAA75633
ID AAA75633 standard; DNA; 38506 BP.
XX
AC AAA75633;
XX
DT 22-JAN-2001 (first entry)
XX
DE Nucleotide sequence of the insert DNA in cosmid pKOS023-27.
XX
KW Narbonolide synthase; polyketide synthase gene; narbonolide polyketide;
KW antibiotic; C12-hydroxyalse; picK; desosamine biosynthesis;
KW desosaminyl transferase enzyme; ketolide; beta-glucosidase enzyme;
KW picromycin biosynthesis; ss.
XX
OS Streptomyces venezuelae.
XX
PN US6117659-A.
XX
PD 12-SEP-2000.
XX
PF 27-MAY-1999; 99US-0320878.
XX
PR 28-MAY-1998; 98US-0087080.
PR 22-SEP-1998; 98US-0100880.
PR 08-FEB-1999; 99US-0119139.
PR 20-MAY-1999; 99US-0134990.
PR 30-APR-1997; 97US-0846247.
PR 06-MAY-1998; 98US-0073538.
PR 28-AUG-1998; 98US-0141908.
XX
PA (KOSA-) KOSAN BIOSCIENCES INC.
XX
PI Ashley G, Betlach MC, Betlach M, Tang L, McDaniel R;
XX
DR WPI; 2000-610844/58.
XX
PT New recombinant picK hydroxylase gene of Streptomyces venezuelae useful
PT for converting ketolides to antibiotics and as antibiotics and
PT intermediates in the synthesis of compounds with pharmaceutical value
PT -
XX
PS Disclosure; Columns 15-32; 117pp; English.
XX
CC The present sequence is used to produce the recombinant DNA compounds
CC of the invention. The specification describes a recombinant DNA compound
CC expressing recombinant polyketide synthase genes in host cells for the
CC production of narbonolide, narbonolide derivatives and polyketides that
CC are useful as antibiotics and as intermediates in the synthesis of
CC compounds with pharmaceutical value. The DNA compounds may also encode
CC a C12-hydroxyalse (picK), desosamine biosynthesis and desosaminyl
CC transferase enzymes (useful for conversion of ketolides to antibiotics),
CC and the beta-glucosidase enzyme (involved in picromycin biosynthesis).
CC These compounds are also useful for increasing the antibiotic activity
CC of a compound relative to the unhydroxylated compound. The recombinant
CC host cells are useful as genetic systems that allow rapid engineering
CC of the narbonolide polyketide synthase. These would be valuable for
CC creating novel ketolide analogs for pharmaceutical applications.
XX
SQ Sequence 38506 BP; 4914 A; 15118 C; 13444 G; 5030 T; 0 other;

Query Match 30.5%; Score 133.6; DB 21; Length 38506;
Best Local Similarity 58.2%; Pred. No. 7.1e-21;
Matches 235; Conservative 0; Mismatches 169; Indels 0; Gaps 0;

```
Qy          1 TACTGGCTCGAAAGCACACAGCCCGGTGCCGGCAACGTGTCAGCAGCCGGACTCGACCCC 60
              |||||||   |   |   |   |||  |||||  || |  |  | ||||| ||||   |
Db      21045 TACTGGCCGCAGCCCGACCTCTCCGCCGCCGGTGACATCACCTCCGCCGGTCTCGGGGCG 21104

Qy         61 ACCGAACACCCCCTACTCGGCGCCACATTGGAACTGGCGACTGACGGTGGAGCGCTTCTT 120
               ||||  |||||  || ||||||||  |  |||  ||  |||     ||  ||    || ||
Db      21105 GCCGAGCACCCGCTGCTCGGCGCGGCCGTGGCGCTCGCGGACTCCGACGGCTGCCTGCTC 21164

Qy        121 GCAGGGCGCTTGTCTTTGAGGTCGCATCCGTGGCTGGCTGACCATGCCGTCGGCGGCACG 180
              | ||| || | ||  |  |  |||| || |||||||| ||||| || || | ||||||
Db      21165 ACGGGGAGCCTCTCCCTCCGTACGCACCCCTGGCTGGCGGACCACGCGGTGGCCGGCACC 21224

Qy        181 GTGCTGCTGTCGGGCGCCACCTTCCTCGAACTCGCCCTTCATGCGGGCACATACGTGGGC 240
              |||||||| ||||  |  | ||| | || || ||  | |  || ||      | || ||
Db      21225 GTGCTGCTGCCGGGAACGGCGTTCGTGGAGCTGGCGTTCCGAGCCGGGGACCAGGTCGGT 21284

Qy        241 TGCGACCGAGTGGATGAGCTGACGCTGCATGCGCCGCTGGTGGTTCCTGTGGATGGGGGT 300
              |||||  |  || || ||||| || ||  | |||||||| ||| | ||      ||| |
Db      21285 TGCGATCTGGTCGAGGAGCTCACCCTCGACGCGCCGCTCGTGCTGCCCCGTCGTGGCGCG 21344

Qy        301 GTGAGTGTGCAGGTTGGGGTTGCGGCTGCGGATGGGGAGGGGCGGCGTTTGGTGAGTGTG 360
```

```
Db      21345 GTCCGTGTGCAGCTGTCCGTCGGCGCGAGCGACGAGTCCGGGCGTCGTACCTTCGGGCTC 21404

Qy        361 TATGCGCGGGGTGGGAGTGCTTGTGGTGGGGGTGGTGCGTCGGG 404
              || ||||     | |    ||     || | ||    |    | || |
Db      21405 TACGCGCACCCGGAGGACGCGCCGGGCGAGGCGGAGTGGACGCG 21448


RESULT 12
AAZ56001
ID   AAZ56001 standard; DNA; 38506 BP.
XX
AC   AAZ56001;
XX
DT   23-MAR-2000  (first entry)
XX
DE   Recombinant cosmid pKOS023-27 containing S. venezuelae PKS genes.
XX
KW   Narbonolide polyketide synthase; PKS; cosmid pKOS023-27; ketolide;
KW   antibiotic production; narbomycin; picromycin; ds.
XX
OS   Streptomyces venezuelae.
XX
FH   Key             Location/Qualifiers
FT   CDS             70..13725
FT                   /*tag= a
FT                   /product= PICAI
FT                   /note= "Narbonolide synthase subunit 1"
FT   CDS             13830..25049
FT                   /*tag= b
FT                   /product= PICAII
FT                   /note= "Narbonolide synthase subunit 2"
FT   CDS             25133..29821
FT                   /*tag= c
FT                   /product= PICAIII
FT                   /note= "Narbonolide synthase subunit 3"
FT   CDS             29924..33964
FT                   /*tag= d
FT                   /product= PICAIV
FT                   /note= "Narbonolide synthase subunit 4"
FT   CDS             33961..34806
FT                   /*tag= e
FT                   /product= PicB
FT                   /note= "Contains typeII thioesterase domain"
FT   CDS             34863..36011
FT                   /*tag= f
FT                   /product= PICCII
FT                   /note= "4-keto-6-deoxyglucose isomerase"
FT   CDS             36159..37439
FT                   /*tag= g
FT                   /product= PICCIII
FT                   /note= "Desosaminyl transferase"
FT   CDS             37529..38242
FT                   /*tag= h
FT                   /product= PICCVI
FT                   /note= "3-amino dimethyltransferase"
XX
PN   WO9961599-A2.
XX
PD   02-DEC-1999.
XX
PF   27-MAY-1999;   99WO-US11814.
XX
PR   28-MAY-1998;   98US-0087080.
PR   28-AUG-1998;   98US-0141908.
PR   22-SEP-1998;   98US-0100880.
PR   08-FEB-1999;   99US-0119139.
XX
PA   (KOSA-) KOSAN BIOSCIENCES INC.
XX
PI   Ashley G,  Betlach MC,  Betlach M,  McDaniel R,  Tang L;
XX
DR   WPI; 2000-072618/06.
DR   P-PSDB; AAY67201, AAY67202, AAY67203, AAY67204, AAY67205, AAY67207,
DR   AAY67208, AAY67211.
XX
PT   New recombinant DNA encoding a domain of narbonolide polyketide
PT   synthase, for production of ketolide antibiotics -
XX
PS   Example 2; Page 16-27; 98pp; English.
XX
CC   This is the recombinant cosmid pKOS023-27 DNA sequence which contains a
CC   Streptomyces venezuelae DNA insert. The cosmid contains open reading
CC   frames which encode the various modules of the narbonolide polyketide
CC   synthase (PKS). The invention relates to recombinant DNA containing a
CC   coding sequence for a narbonolide PKS. Polyketides are compounds
CC   synthesised from 2-carbon units through a series of condensations and
CC   subsequent modifications. Modular PKSs are responsible for the production
CC   of many antibiotics including picromycin. The narbonolide PKS consists of
CC   a loading module, six extender modules, and two thioester domains. Four
CC   proteins make up the narbonolide PKS (PICAI, PICAII, PICAIII and PICAIV).
CC   PICAI includes the loading module and extender modules 1 and 2, PICAII
CC   includes extender modules 3 and 4, PICAIII includes extender module 5 and
CC   PICAIV includes extender module 6 and a type II thioesterase domain. The
CC   second type II thioesterase domain is found on the PICB protein. The
CC   nucleotide sequences encoding all of these proteins can be isolated in
CC   recombinant form from the recombinant cosmid pKOS023-27. Narbonolide is
CC   desosaminylated in S. venezuelae to yield narbomycin, the desosaminyl
CC   transferase enzyme is required for this  conversion, and the desosamine
CC   biosynthetic genes are also found in cosmid pKOS023-27. The recombinant
CC   DNA of the invention is used to express, in transformed cells,
CC   narbonolide (or its derivatives) or other ketolides (particularly
CC   hybrids), which may then be converted (e.g. by other enzymes
CC   recombinantly expressed in the same hosts) to polyketide antibiotics or
CC   their intermediates. The antibiotics are useful in human or veterinary
CC   medicine.
XX
SQ   Sequence 38506 BP; 4914 A; 15118 C; 13444 G; 5030 T; 0 other;

  Query Match             30.5%;  Score 133.6;  DB 21;  Length 38506;
  Best Local Similarity   58.2%;  Pred. No. 7.1e-21;
  Matches  235;  Conservative    0;  Mismatches  169;  Indels    0;  Gaps    0;

Qy          1 TACTGGCTCGAAAGCACACAGCCCGGTGCCGGCAACGTGTCAGCAGCCGGACTCGACCCC 60
              |||||||    |   |   |   |||  |||||  || |  |  | ||||| ||||   |
Db      21045 TACTGGCCGCAGCCCGACCTCTCCGCCGCCGGTGACATCACCTCCGCCGGTCTCGGGGCG 21104

Qy         61 ACCGAACACCCCCTACTCGGCGCCACATTGGAACTGGCGACTGACGGTGGAGCGCTTCTT 120
                |||| ||||| || ||||||||  |  |||  || |||  ·  ||  ||    || ||
Db      21105 GCCGAGCACCCGCTGCTCGGCGCGGCCGTGGCGCTCGCGGACTCCGACGGCTGCCTGCTC 21164

Qy        121 GCAGGGCGCTTGTCTTTGAGGTCGCATCCGTGGCTGGCTGACCATGCCGTCGGCGGCACG 180
              | ||| || | ||  |  |  |||| || |||||||| ||||| || || | ||||||
Db      21165 ACGGGGAGCCTCTCCCTCCGTACGCACCCCTGGCTGGCGGACCACGCGGTGGCCGGCACC 21224

Qy        181 GTGCTGCTGTCGGGCGCCACCTTCCTCGAACTCGCCCTTCATGCGGGCACATACGTGGGC 240
              |||||||| ||||  |  | ||| | || || ||  | |  || ||     | || ||
Db      21225 GTGCTGCTGCCGGGAACGGCGTTCGTGGAGCTGGCGTTCCGAGCCGGGGACCAGGTCGGT 21284

Qy        241 TGCGACCGAGTGGATGAGCTGACGCTGCATGCGCCGCTGGTGGTTCCTGTGGATGGGGGT 300
              |||||  |  || || ||||| || ||  | |||||||| ||| | ||     ||| |
Db      21285 TGCGATCTGGTCGAGGAGCTCACCCTCGACGCGCCGCTCGTGCTGCCCCGTCGTGGCGCG 21344

Qy        301 GTGAGTGTGCAGGTTGGGGTTGCGGCTGCGGATGGGGAGGGGCGGCGTTTGGTGAGTGTG 360
              ||  |||||||| |    || |  ||    || | |   ||||| |||    |  | |
Db      21345 GTCCGTGTGCAGCTGTCCGTCGGCGCGAGCGACGAGTCCGGGCGTCGTACCTTCGGGCTC 21404

Qy        361 TATGCGCGGGGTGGGAGTGCTTGTGGTGGGGGTGGTGCGTCGGG 404
              || ||||     | |    ||     || | ||    |    | || |
Db      21405 TACGCGCACCCGGAGGACGCGCCGGGCGAGGCGGAGTGGACGCG 21448


RESULT 13
AAD17184
ID   AAD17184 standard; DNA; 65140 BP.
XX
```

```
AC   AAD17184;
XX
DT   29-NOV-2001  (first entry)
XX
DE   Streptomyces noursei nys1 DNA of nystatin PKS gene cluster.
XX
KW   Polyketide synthase; PKS; macrolide; nystatin; PKS gene cluster;
KW   antifungal; antibiotic; nys1; ds.
XX
OS   Streptomyces noursei.
XX
FH   Key             Location/Qualifiers
FT   CDS             complement (1..1035)
FT                   /*tag=  a
FT                   /product= "NysD2 partial protein"
FT                   /note= "CDS does not include stop codon"
FT   CDS             complement (1056..2576)
FT                   /*tag=  b
FT                   /product= "NysD1 protein"
FT   CDS             2806..6906
FT                   /*tag=  c
FT                   /product= "NysA protein"
FT   CDS             6952..16530
FT                   /*tag=  d
FT                   /product= "NysB protein"
FT   CDS             16550..49840
FT                   /*tag=  e
FT                   /product= "NysC protein"
FT   CDS             50260..51015
FT                   /*tag=  f
FT                   /product= "NysE protein"
FT   CDS             51405..54305
FT                   /*tag=  g
FT                   /product= "NysR1 protein"
FT   CDS             54329..57190
FT                   /*tag=  h
FT                   /product= "NysR2 protein"
FT                   /note= "CDS does not include start codon"
FT   CDS             57180..59963
FT                   /*tag=  i
FT                   /product= "NysR3 protein"
FT   CDS             60415..61047
FT                   /*tag=  j
FT                   /product= "NysR4 (short) protein"
FT                   /note= "CDS does not include start codon"
FT   CDS             61736..62497
FT                   /*tag=  k
FT                   /product= "NysR5 protein"
FT                   /note= "CDS does not include start codon"
FT   CDS             complement (62551..63615)
FT                   /*tag=  l
FT                   /product= "ORF2 protein"
FT                   /note= "CDS does not include start codon"
FT   CDS             63765..64961
FT                   /*tag=  m
FT                   /product= "ORF1 protein"
XX
PN   WO200159126-A2.
XX
PD   16-AUG-2001.
XX
PF   08-FEB-2001; 2001WO-GB00509.
XX
PR   08-FEB-2000; 2000GB-0002840.
PR   10-APR-2000; 2000GB-0008786.
PR   14-APR-2000; 2000GB-0009387.
XX
PA   (UYNO-) UNIV NORGES TEKNISK NATURVITENSKAPELIGE.
PA   (SNTF ) SINTEF STIFTELSEN IND TEK FORSK.
PA   (ALPH-) ALPHARMA AS.
PA   (SINV-) SINVENT AS.
PA   (DZIE/) DZIEGLEWSKA H.
PA   (ZOTC/) ZOTCHEV S B.
PA   (SEKU/) SEKUROVA O N.
PA   (FJAE/) FJAERVIK E.
PA   (BRAU/) BRAUTASET T.
PA   (STRO/) STROM A R.
XX
PI   Zotchev SB,  Sekurova ON,  Fjaervik E,  Brautaset T,  Strom AR;
PI   Valla S,  Ellingsen TE,  Sletta H,  Gulliksen O;
XX
DR   WPI; 2001-557614/62.
DR   P-PSDB; AAE10125, AAE10126, AAE10127, AAE10128, AAE10129, AAE10130,
DR   AAE10131, AAE10132, AAE10133, AAE10134, AAE10135, AAE10136, AAE10137.
XX
PT   New nystatin polyketide synthase polynucleotides and polypeptides,
PT   useful as antibiotics and antifungals -
XX
PS   Claim 2; Page 116-151; 266pp; English.
XX
CC   The present invention relates to the cloning and sequencing of the gene
CC   cluster encoding a modular type I polyketide synthase (PKS) enzyme
CC   involved in the biosynthesis of the macrolide antibiotic nystatin.
CC   The nystatin PKS is useful as antifungal antibiotics. The present
CC   sequence is a Streptomyces noursei nys1 DNA of nystatin PKS gene cluster.
XX
SQ   Sequence 65140 BP; 8270 A; 25171 C; 22273 G; 9426 T; 0 other;

  Query Match             30.3%;  Score 132.8;  DB 22;  Length 65140;
  Best Local Similarity   60.1%;  Pred. No. 1.1e-20;
  Matches  221;  Conservative    0;  Mismatches  147;  Indels    0;  Gaps    0;

Qy          1 TACTGGCTCGAAAGCACACAGCCCGGTGCCGGCAACGTGTCAGCAGCCGGACTCGACCCC 60
              | ||||| ||    |  ||   || |  ||| | ||||    || ||||| || | | ||
Db      24428 TTCTGGCCCGCCGCCCGACCCGCCCGCCCCGACGACGTCCGCGCCGCCGGCCTGGGCGCC 24487

Qy         61 ACCGAACACCCCCTACTCGGCGCCACATTGGAACTGGCGACTGACGGTGGAGCGCTTCTT 120
              ||||||||||||| ||||||||| |  || ||||  |    ||||  || || ||  |
Db      24488 GCCGAACACCCCCTCCTCGGCGCCGCCGTGCAACTCCCCGACGACGACGGCGCACTCTTC 24547

Qy        121 GCAGGGCGCTTGTCTTTGAGGTCGCATCCGTGGCTGGCTGACCATGCCGTCGGCGGCACG 180
              | || ||| | ||  || |  | || ||||||||||| |||||  |||||   |||||
Db      24548 ACCGGCCGCCTCTCCCTGCGCACCCACCCGTGGCTGGCCGACCACACCGTCCTGGGCACC 24607

Qy        181 GTGCTGCTGTCGGGCGCCACCTTCCTCGAACTCGCCCTTCATGCGGGCACATACGTGGGC 240
              || |||||  ||||| || |  |  | |||||||| | |  ||||||    |    |||
Db      24608 GTCCTGCTCCCGGGCACCGCACTGGTGGAACTCGCCGTCCGCGCGGGCGACGAGACCGGC 24667

Qy        241 TGCGACCGAGTGGATGAGCTGACGCTGCATGCGCCGCTGGTGGTTCCTGTGGATGGGGGT 300
              ||| ||   | || || || || ||    ||||| |||    | || | ||| || |
Db      24668 AGCGGCCACCTCGAAGAACTCACCCTCGCCGCGCCCCTGACCCTCCCCGAGGACGGCGCC 24727

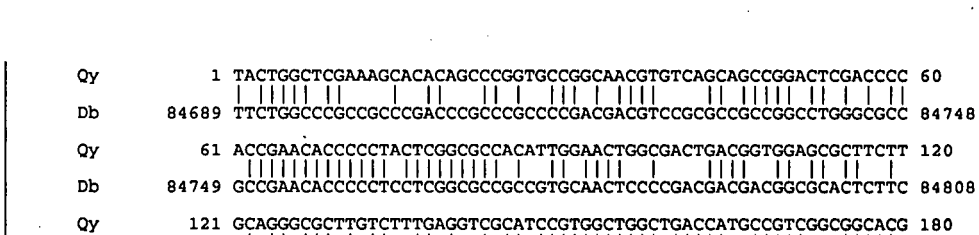
Qy        301 GTGAGTGTGCAGGTTGGGGTTGCGGCTGCGGATGGGGAGGGGCGGCGTTTGGTGAGTGTG 360
                | |||||  | || |   | || || |      || || ||    || |  ||
Db      24728 ACCCTCCTCCAGGTCCGCGTCGGATCCGCCGACGACACCGGCCGCCGCACCGTCACCGTC 24787

Qy        361 TATGCGCG 368
              | || ||
Db      24788 CACGCCCG 24795


RESULT 14
AAD17186
ID   AAD17186 standard; DNA; 125401 BP.
XX
AC   AAD17186;
XX
DT   29-NOV-2001  (first entry)
XX
DE   Streptomyces noursei nystatin PKS gene cluster DNA.
XX
KW   Polyketide synthase; PKS; macrolide; nystatin; PKS gene cluster;
KW   antifungal; antibiotic; ds.
XX
OS   Streptomyces noursei.
XX
```

```
FH   Key             Location/Qualifiers
FT   CDS             6337..34771
FT                   /*tag= a
FT                   /product= "NysI complete protein"
FT   CDS             34792..51099
FT                   /*tag= b
FT                   /product= "NysJ protein"
FT   CDS             51155..57355
FT                   /*tag= c
FT                   /product= "NysK protein"
FT   CDS             57503..58687
FT                   /*tag= d
FT                   /product= "NysL protein"
FT   CDS             complement (58786..58980)
FT                   /*tag= e
FT                   /product= "NysM protein"
FT                   /note= "CDS does not include start codon"
FT   CDS             complement (59045..60241)
FT                   /*tag= f
FT                   /product= "NysN protein"
FT                   /note= "CDS does not include start codon"
FT   CDS             complement (60238..61296)
FT                   /*tag= g
FT                   /product= "NysD2 complete protein"
FT   CDS             120628..121308
FT                   /*tag= h
FT                   /product= "NysR4 (long) protein"
XX
PN   WO200159126-A2.
XX
PD   16-AUG-2001.
XX
PF   08-FEB-2001; 2001WO-GB00509.
XX
PR   08-FEB-2000; 2000GB-0002840.
PR   10-APR-2000; 2000GB-0008786.
PR   14-APR-2000; 2000GB-0009387.
XX
PA   (UYNO-) UNIV NORGES TEKNISK NATURVITENSKAPELIGE.
PA   (SNTF ) SINTEF STIFTELSEN IND TEK FORSK.
PA   (ALPH-) ALPHARMA AS.
PA   (SINV-) SINVENT AS.
PA   (DZIE/) DZIEGLEWSKA H.
PA   (ZOTC/) ZOTCHEV S B.
PA   (SEKU/) SEKUROVA O N.
PA   (FJAE/) FJAERVIK E.
PA   (BRAU/) BRAUTASET T.
PA   (STRO/) STROM A R.
XX
PI   Zotchev SB,  Sekurova ON,  Fjaervik E,  Brautaset T,  Strom AR;
PI   Valla S,  Ellingsen TE,  Sletta H,  Gulliksen O;
XX
DR   WPI; 2001-557614/62.
DR   P-PSDB; AAE10143, AAE10144, AAE10145, AAE10146, AAE10147, AAE10148,
DR   AAE10149, AAE10150.
XX
PT   New nystatin polyketide synthase polynucleotides and polypeptides,
PT   useful as antibiotics and antifungals -
XX
PS   Claim 1; Page 188-254; 266pp; English.
XX
CC   The present invention relates to the cloning and sequencing of the gene
CC   cluster encoding a modular type I polyketide synthase (PKS) enzyme
CC   involved in the biosynthesis of the macrolide antibiotic nystatin.
CC   The nystatin PKS is useful as antifungal antibiotics. The present
CC   sequence is a Streptomyces noursei nystatin PKS gene cluster DNA.
XX
SQ   Sequence 125401 BP; 15664 A; 49692 C; 42871 G; 17174 T; 0 other;

  Query Match           30.3%;  Score 132.8;  DB 22;  Length 125401;
  Best Local Similarity  60.1%;  Pred. No. 1.2e-20;
  Matches  221;  Conservative    0;  Mismatches  147;  Indels    0;  Gaps    0;
```



```
Qy          1 TACTGGCTCGAAAGCACACAGCCCGGTGCCGGCAACGTGTCAGCAGCCGGACTCGACCCC 60
              |||||| ||      |  ||   || |  ||| | ||||    || ||||| || | | ||
Db      84689 TTCTGGCCCGCCGCCCGACCCGCCCGCCCCGACGACGTCCGCGCCGCCGGCCTGGGCGCC 84748

Qy         61 ACCGAACACCCCCTACTCGGCGCCACATTGGAACTGGCGACTGACGGTGGAGCGCTTCTT 120
              |||||||||||||| |||||||||  |  || ||||  |    ||||  || || ||  |
Db      84749 GCCGAACACCCCCTCCTCGGCGCCGCCGTGCAACTCCCCGACGACGACGGCGCACTCTTC 84808

Qy        121 GCAGGGCGCTTGTCTTTGAGGTCGCATCCGTGGCTGGCTGACCATGCCGTCGGCGGCACG 180
              | || ||| | ||  || |  | || ||||||||||| |||||  |||||    |||||
Db      84809 ACCGGCCGCCTCTCCCTGCGCACCCACCCGTGGCTGGCCGACCACACCGTCCTGGGCACC 84868

Qy        181 GTGCTGCTGTCGGGCGCCACCTTCCTCGAACTCGCCCTTCATGCGGGCACATACGTGGGC 240
              || |||||  ||||| || |  |  | ||||||||| | |  ||||||     |    |||
Db      84869 GTCCTGCTCCCGGGCACCGCACTGGTGGAACTCGCCGTCCGCGCGGGCGACGAGACCGGC 84928

Qy        241 TGCGACCGAGTGGATGAGCTGACGCTGCATGCGCCGCTGGTGGTTCCTGTGGATGGGGGT 300
              ||| ||   | || || || || ||     ||||| |||    | || | ||| || |
Db      84929 AGCGGCCACCTCGAAGAACTCACCCTCGCCGCGCCCCTGACCCTCCCCGAGGACGGCGCC 84988

Qy        301 GTGAGTGTGCAGGTTGGGGTTGCGGCTGCGGATGGGGAGGGGCGGCGTTTGGTGAGTGTG 360
              |  |||||  | || |   | || || |     || || ||    || |  ||
Db      84989 ACCCTCCTCCAGGTCCGCGTCGGATCCGCCGACGACACCGGCCGCCGCACCGTCACCGTC 85048

Qy        361 TATGCGCG 368
              | || ||
Db      85049 CACGCCCG 85056
```

```
RESULT 15
AAD17185
ID   AAD17185 standard; DNA; 27541 BP.
XX
AC   AAD17185;
XX
DT   29-NOV-2001  (first entry)
XX
DE   Streptomyces noursei nys2 DNA of nystatin PKS gene cluster.
XX
KW   Polyketide synthase; PKS; macrolide; nystatin; PKS gene cluster;
KW   antifungal; antibiotic; nys2; ds.
XX
OS   Streptomyces noursei.
XX
FH   Key             Location/Qualifiers
FT   CDS             complement (454..1191)
FT                   /*tag= a
FT                   /product= "NysF protein"
FT                   /note= "CDS does not include start codon"
FT   CDS             complement (1275..3092)
FT                   /*tag= b
FT                   /product= "NysG protein"
FT   CDS             complement (3070..4824)
FT                   /*tag= c
FT                   /note= "CDS does not include start codon"
FT                   /product= "NysH protein"
FT   CDS             5122..6156
FT                   /*tag= d
FT                   /product= "NysD3 protein"
FT   CDS             6338..27541
FT                   /*tag= e
FT                   /product= "NysI partial protein"
FT                   /note= "CDS does not include stop codon"
XX
PN   WO200159126-A2.
XX
PD   16-AUG-2001.
XX
PF   08-FEB-2001; 2001WO-GB00509.
XX
PR   08-FEB-2000; 2000GB-0002840.
PR   10-APR-2000; 2000GB-0008786.
PR   14-APR-2000; 2000GB-0009387.
```

XX
PA (UYNO-) UNIV NORGES TEKNISK NATURVITENSKAPELIGE.
PA (SNTF) SINTEF STIFTELSEN IND TEK FORSK.
PA (ALPH-) ALPHARMA AS.
PA (SINV-) SINVENT AS.
PA (DZIE/) DZIEGLEWSKA H.
PA (ZOTC/) ZOTCHEV S B.
PA (SEKU/) SEKUROVA O N.
PA (FJAE/) FJAERVIK E.
PA (BRAU/) BRAUTASET T.
PA (STRO/) STROM A R.
XX
PI Zotchev SB, Sekurova ON, Fjaervik E, Brautaset T, Strom AR;
PI Valla S, Ellingsen TE, Sletta H, Gulliksen O;
XX
DR WPI; 2001-557614/62.
DR P-PSDB; AAE10138, AAE10139, AAE10140, AAE10141, AAE10142.
XX
PT New nystatin polyketide synthase polynucleotides and polypeptides,
PT useful as antibiotics and antifungals -
XX
PS Claim 2; Page 151-166; 266pp; English.
XX
CC The present invention relates to the cloning and sequencing of the gene
CC cluster encoding a modular type I polyketide synthase (PKS) enzyme
CC involved in the biosynthesis of the macrolide antibiotic nystatin.
CC The nystatin PKS is useful as antifungal antibiotics. The present
CC sequence is a Streptomyces noursei nys2 DNA of nystatin PKS gene cluster.
XX
SQ Sequence 27541 BP; 3517 A; 10766 C; 9529 G; 3729 T; 0 other;

Query Match 30.0%; Score 131.2; DB 22; Length 27541;
Best Local Similarity 59.8%; Pred. No. 2.4e-20;
Matches 220; Conservative 0; Mismatches 148; Indels 0; Gaps 0;

```
Qy          1 TACTGGCTCGAAAGCACACAGCCCGGTGCCGGCAACGTGTCAGCAGCCGGACTCGACCCC 60
              ||||||| | |   |   |||  |||  ||| | || ||    | | ||| ||||   ||
Db       9077 TACTGGCCCAAGGCCCTCCAGAGCGGCACCGCCGACCTGCGCTCGGTCGGCCTCGGTGCC 9136

Qy         61 ACCGAACACCCCCTACTCGGCGCCACATTGGAACTGGCGACTGACGGTGGAGCGCTTCTT 120
               || | ||||| || |||  |||| | |     || ||    |  || ||  | || ||
Db       9137 GCCCACCACCCGCTGCTCTCCGCCGCCGTCTCCCTCGCCGACGCAGGCGGCACCCTGCTC 9196

Qy        121 GCAGGGCGCTTGTCTTTGAGGTCGCATCCGTGGCTGGCTGACCATGCCGTCGGCGGCACG 180
              | || ||| | ||   |  | | || || ||||| || |||||  ||||| |||||||
Db       9197 ACCGGCCGCCTCTCCCGGCAGACCCACCCCTGGCTCGCCGACCACACCGTCCGCGGCACC 9256

Qy        181 GTGCTGCTGTCGGGCGCCACCTTCCTCGAACTCGCCCTTCATGCGGGCACATACGTGGGC 240
                |||||| | ||  || ||||||||||||||||| | |  || |||    | || |||
Db       9257 ACCCTGCTGCCCGGTACCGCCTTCCTCGAACTCGCCGTCCGCGCCGGCGACGAGGTCGGC 9316

Qy        241 TGCGACCGAGTGGATGAGCTGACGCTGCATGCGCCGCTGGTGGTTCCTGTGGATGGGGGT 300
              |||||||| || || || || || ||     || |||||  || | || |    | || ||
Db       9317 TGCGACCGCGTCGAGGAACTCACCCTCGCCGCACCGCTCCTGCTGCCCGAACAGGGCGGC 9376

Qy        301 GTGAGTGTGCAGGTTGGGGTTGCGGCTGCGGATGGGGAGGGGCGGCGTTTGGTGAGTGTG 360
              ||    || ||| |  || | |      | || | |   || || ||    || |  ||
Db       9377 GTCCAGGTCCAGTTGTGGATCGGCAACCCCGACGTGTCCGGTCGCCGCACCGTCAACGTC 9436

Qy        361 TATGCGCG 368
               |.|| ||
Db       9437 CACGCCCG 9444
```

Search completed: June 17, 2003, 13:48:47
Job time : 93.2209 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 17, 2003, 11:26:45 ; Search time 561.472 Seconds
(without alignments)
12633.968 Million cell updates/sec

Title: US-09-914-286-1_COPY_8947_9384
Perfect score: 438
Sequence: 1 tactggctcgaaagcacaca.....gtcatgcctcgggggtgctg 438

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	67.2	15.3	384	17	CNS035G7	AL228688 Tetraodon
	2	65.8	15.0	811	13	BI888042	BI888042 ZF637-1-0
c	3	64.8	14.8	675	17	AG139676	AG139676 Pan trogl
	4	64.4	14.7	925	17	CNS0091P	AL053013 Drosophil
c	5	64.2	14.7	1101	17	CNS016TH	AL107183 Drosophil
c	6	64	14.6	796	17	AG076826	AG076826 Pan trogl
	7	64	14.6	985	14	BQ068225	BQ068225 AGENCOURT
	8	63	14.4	922	17	CNS0073W	AL066784 Drosophil
c	9	63	14.4	1094	17	CNS008N8	AL052088 Drosophil
c	10	62.8	14.3	1168	13	BM468388	BM468388 AGENCOURT
	11	62.6	14.3	797	17	AZ191837	AZ191837 SP_1020_A
	12	62.4	14.2	884	17	CNS006U0	AL065923 Drosophil
c	13	62	14.2	886	14	BQ433302	BQ433302 AGENCOURT
c	14	61.6	14.1	1101	17	CNS00397	AL063912 Drosophil
c	15	61.4	14.0	1062	14	BQ653252	BQ653252 AGENCOURT
	16	61.2	14.0	857	17	AG044313	AG044313 Pan trogl
c	17	61	13.9	638	17	CNS046XY	AL277279 Tetraodon
	18	61	13.9	842	17	CNS02J8H	AL199898 Tetraodon
c	19	61	13.9	969	14	BQ712788	BQ712788 AGENCOURT
	20	61	13.9	969	17	CNS03PO0	AL254889 Tetraodon
	21	60.8	13.9	378	9	AL535833	AL535833 AL535833
	22	60.8	13.9	1001	13	BM549244	BM549244 AGENCOURT
c	23	60.8	13.9	1004	17	CNS015QM	AL105784 Drosophil
c	24	60.6	13.8	586	17	CNS01UD1	AL167662 Tetraodon
	25	60.6	13.8	591	17	AQ051635	AQ051635 nbxb0002c
	26	60.6	13.8	897	14	BQ690564	BQ690564 AGENCOURT
	27	60.6	13.8	992	17	CNS04R5Q	AL303479 Tetraodon
c	28	60.6	13.8	1139	12	BF797208	BF797208 602257701
	29	60.6	13.8	1461	13	BM474555	BM474555 AGENCOURT
	30	60.4	13.8	583	12	BG786318	BG786318 SEAUMC006
c	31	60.4	13.8	752	17	AG135365	AG135365 Pan trogl
	32	60.4	13.8	901	17	AG043735	AG043735 Pan trogl
c	33	60.4	13.8	1109	13	BM543244	BM543244 AGENCOURT
c	34	60.2	13.7	327	17	CNS037PM	AL231619 Tetraodon
	35	60.2	13.7	1059	17	CNS0155J	AL105025 Drosophil
c	36	60.2	13.7	1245	14	BM908713	BM908713 AGENCOURT
c	37	60	13.7	449	17	BH177718	BH177718 010_G_04-
c	38	60	13.7	449	17	CNS07K77	AL614677 T7 end of
c	39	60	13.7	793	17	AG141003	AG141003 Pan trogl
	40	60	13.7	812	9	AL537183	AL537183 AL537183
	41	60	13.7	847	12	BG784057	BG784057 SEAUMC004
	42	60	13.7	1004	17	AZ671323	AZ671323 ENTKO33TR
	43	60	13.7	1020	17	AZ679537	AZ679537 ENTKU45TR
c	44	59.8	13.7	324	17	CNS00ICM	AL074510 Drosophil
	45	59.8	13.7	880	17	CNS04463	AL273684 Tetraodon

ALIGNMENTS

RESULT 1
CNS035G7
LOCUS CNS035G7 384 bp DNA linear GSS 15-MAY-2000
DEFINITION Tetraodon nigroviridis genome survey sequence T7 end of clone 213J04 of library G from Tetraodon nigroviridis, genomic survey sequence.
ACCESSION AL228688
VERSION AL228688.1 GI:7887678
KEYWORDS GSS; genome survey sequence.
SOURCE Tetraodon nigroviridis.
ORGANISM Tetraodon nigroviridis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes; Tetraodontidae; Tetraodon.
REFERENCE 1 (bases 1 to 384)
AUTHORS Roest-Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C., Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F., Saurin,W. and Weissenbach,J.
TITLE Human gene number estimate provided by genome wide analysis using Tetraodon nigroviridis DNA sequence
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 384)
AUTHORS Roest-Crollius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C., Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and Weissenbach,J.
TITLE Charaterization and repeat analysis of the compact genome of the freshwater pufferfish Tetraodon nigroviridis
JOURNAL Unpublished

```
REFERENCE   3  (bases 1 to 384)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-APR-2000)
COMMENT     This sequence is a single read and was generated as part of a large
            scale clone-end sequencing project of the Tetraodon nigroviridis
            genome. For more information, please take a look at
            http://www.genoscope.cns.fr/Tetraodon.
FEATURES             Location/Qualifiers
     source          1. .384
                     /organism="Tetraodon nigroviridis"
                     /db_xref="taxon:99883"
                     /clone="213J04"
                     /clone_lib="G"
                     /note="Genoscope sequence ID : C0AG213DE02LP1~end : T7"
BASE COUNT       23 a     39 c    199 g     54 t     69 others
ORIGIN

  Query Match             15.3%;  Score 67.2;  DB 17;  Length 384;
  Best Local Similarity   39.8%;  Pred. No. 2.6e-05;
  Matches  103;  Conservative   48;  Mismatches  108;  Indels    0;  Gaps    0;

Qy        177 CACGGTGCTGCTGTCGGGCGCCACCTTCCTCGAACTCGCCCTTCATGCGGGCACATACGT 236
              | |   |:  ::|:||   : :   ::  :|::  :|: : ||   |:|:::    | :|
Db         91 CTCTCTSTSBSTSTCTCTSCSKSKSKSTSTSKSTSTSTSTCTSTCTSCSKSTSTSTSTST 150

Qy        237 GGGCTGCGACCGAGTGGATGAGCTGACGCTGCATGCGCCGCTGGTGGTTCCTGTGGATGG 296
              : :| |::   : | :|  : : :|  : :|  :| | : :|: ||:    |:|: :|:
Db        151 STSCSGSKSGGSGGSKGGGKGKGKGKGKGTGKGGGKGKGKGKGGGKGKGGGKGKGKGK 210

Qy        297 GGGTGTGAGTGTGCAGGTTGGGGTTGCGGCTGCGGATGGGGAGGGGCGGCGTTTGGTGAG 356
              |:|:||| |:| |  :|: ||||  | ||  | ||  |||| :||| |: | : || | |
Db        211 GKGKGTGKGKGGGGGKGKGGGGGGGGGGGGGGGGGGGGGGGGGTKGGGGKGGGKGGGGGGG 270

Qy        357 TGTGTATGCGCGGGGTGGGAGTGCTTGTGGTGGGGGTGGTGCGTCGGGTGGGGTGTGGAC 416
              | |   | | |||| ||| | |   | || ||||| || | |: ||| |||| | ||
Db        271 GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGKGKGGGGGGGGGGGGGGG 330

Qy        417 GTGTCATGCCTCGGGGGTG 435
              | |     |    | ||| |
Db        331 GGGGGGGGGGGTGNGGGGG 349


RESULT 2
BI888042
LOCUS       BI888042                811 bp    mRNA    linear   EST 12-OCT-2001
DEFINITION  ZF637-1-002297 Zebrafish shield stage whole embryo cDNA library
            MPMGp637 Danio rerio cDNA clone MPMGp637_10F13;MPMGp637F1310 5',
            mRNA sequence.
ACCESSION   BI888042
VERSION     BI888042.1  GI:16095313
KEYWORDS    EST.
SOURCE      zebrafish.
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes
            ; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 811)
  AUTHORS   Clark,M., Aanstad,P., Hennig,S., Johnson,S.L. and Lehrach,H.
  TITLE     EST sequencing of a zebrafish shield stage cDNA library normalised
            by oligonucleotide fingerprinting
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Hennig S
            laboraty 123, dept.Lehrach
            Max-Planck-Institut fuer Molekulare Genetik
            Ihnestr.63-73, D-14195 Berlin, Germany
            Tel: +49 30 8413 1612
            Fax: +49 30 8413 1380
            Email: hennig@molgen.mpg.de
            5' EST sequencing of clones from a zebrafish shield stage library,
            normalised from 55,000 starting clones by oligonucleotide
            fingerprinting
            High quality sequence stop: 811.
FEATURES             Location/Qualifiers
     source          1. .811
                     /organism="Danio rerio"
                     /db_xref="taxon:7955"
                     /clone="MPMGp637_10F13;MPMGp637F1310"
                     /clone_lib="Zebrafish shield stage whole embryo cDNA
                     library MPMGp637"
                     /tissue_type="whole embryo"
                     /dev_stage="shield stage, 6 hrs post-fertilisation"
                     /lab_host="E.coli, XL1 blue MRF"
                     /note="Vector: pSport1; Site_1: NotI; Site_2: SalI;
                     oligo-dT-NotI primed, SalI adaptors, directionally cloned,
                     library normalised by oligonucleotide fingerprinting"
BASE COUNT       47 a    219 c    519 g     16 t     10 others
ORIGIN

  Query Match             15.0%;  Score 65.8;  DB 13;  Length 811;
  Best Local Similarity   52.3%;  Pred. No. 6.2e-05;
  Matches  145;  Conservative    0;  Mismatches  132;  Indels    0;  Gaps    0;

Qy        143 CGCATCCGTGGCTGGCTGACCATGCCGTCGGCGGCACGGTGCTGCTGTCGGGCGCCACCT 202
              ||| | || ||  |||   ||     | ||| ||  ||| | | | | |  ||| |
Db        226 CGCCTGCGGGGGGGGCCCCCCCCCCGCCGCGGGGGGGCGGGGGGGGGGGGCCCGCGGCGGGC 285

Qy        203 TCCTCGAACTCGCCCTTCATGCGGGCACATACGTGGGCTGCGACCGAGTGGATGAGCTGA 262
              || |   | | |||  |  |||||       || |||  | |  || | | ||  | || |
Db        286 CCCCCCTCCCCCCCCCCCCCGCGGGGCGGGGCGGGGGGGGGGGGGCGGGGGGGGGGGCGGC 345

Qy        263 CGCTGCATGCGCCGCTGGTGGTTCCTGTGGATGGGGGTGTGAGTGTGCAGGTTGGGGTTG 322
              | |   | | |  || ||     | ||  ||||| | |   | |  ||   |||  |
Db        346 GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGCGGGGGGGGGTGGGGGG 405

Qy        323 CGGCTGCGGATGGGGAGGGGCGGCGTTTGGTGAGTGTGTATGCGCGGGGTGGGAGTGCTT 382
              ||  |||| | || |||| ||     || | | | |  ||| |||| ||| | |
Db        406 GGGGCGCGGGGGCGGGGGGGGGGGCGGGGGGGCGGGGGGGGGCGGGGGGGGGGGGGGGGG 465

Qy        383 GTGGTGGGGGTGGTGCGTCGGGTGGGGTGTGGACGTG 419
              | || ||| | || |||  ||| |||| | || || |
Db        466 GGGGGGGGCGGGGGGCGGGGGGCGGGGGGGGGGCGGG 502


RESULT 3
AG139676/c
LOCUS       AG139676                675 bp    DNA     linear   GSS 04-NOV-2001
DEFINITION  Pan troglodytes DNA, clone: PTB-154D18.R, genomic survey sequence.
ACCESSION   AG139676
VERSION     AG139676.1  GI:16669354
KEYWORDS    GSS.
SOURCE      Pan troglodytes male lymphoblast DNA, clone_lib:PTB Chimpanzee Male
            BAC Library clone:PTB-154D18.R.
  ORGANISM  Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE   1
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     BAC end sequences of Library PTB
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 675)
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            (E-mail:chimpbes@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT     Clones are derived from the chimpanzee BAC library PTB This BAC end
            was generated during the R&D process and may have higher chance of
            clone tracking errors.
             PRIMERS
                Sequencing: M13Rev
             LIBRARY
```

```
                Vector    : pKS145
                R.Site 1  : SacI
                R.Site 2  : SacI.
FEATURES            Location/Qualifiers
     source         1. .675
                    /organism="Pan troglodytes"
                    /db_xref="taxon:9598"
                    /clone="PTB-154D18.R"
                    /sex="male"
                    /cell_type="lymphoblast"
                    /clone_lib="PTB Chimpanzee Male BAC Library"
BASE COUNT      130 a    453 c     38 g     51 t      3 others
ORIGIN

  Query Match             14.8%;  Score 64.8;  DB 17;  Length 675;
  Best Local Similarity   59.0%;  Pred. No. 0.0001;
  Matches  111;  Conservative    0;  Mismatches   77;  Indels    0;  Gaps    0;

Qy        248 GAGTGGATGAGCTGACGCTGCATGCGCCGCTGGTGGTTCCTGTGGATGGGGGTGTGAGTG 307
              | | ||  | |   | | ||    |    |  || ||    |||||  |||||| ||| | |
Db        568 GGGGGGTGGGGGGGGGGGTGGGGGGTGGGTGGGGGGGGGGTGTGGGTGGGGGGGTGGGGG 509

Qy        308 TGCAGGTTGGGGTTGCGGCTGCGGATGGGGAGGGGCGGCGTTTGGTGAGTGTGTATGCGC 367
                |  ||  |||| || || || ||  |||| |||| || || ||| | | | |   | |
Db        508 GGGGGGGGGGGGGTGGGGGTGGGGGGGGGGGGGGGGGGGGTGTGGGGGGGGGGGGGGGGG 449

Qy        368 GGGGTGGGAGTGCTTGTGGTGGGGGTGGTGCGTCGGGTGGGGTGTGGACGTGTCATGCCT 427
              || | ||| |||   | |||||||| || | || ||| ||||   ||  |||  | |
Db        448 GGTGGGGGGGTGGGGGGGGTGGGGGGGGGGGGTGGGGGGGGGGTGGGGGGTGGGAGGTGG 389

Qy        428 CGGGGGTG 435
                |||||
Db        388 GGGGGGGG 381


RESULT 4
CNS0091P
LOCUS       CNS0091P                 925 bp    DNA     linear   GSS 03-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence TET3 end of BAC #
            BACR19D16 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL053013
VERSION     AL053013.1  GI:4934461
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 925)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES            Location/Qualifiers
     source         1. .925
                    /organism="Drosophila melanogaster"
                    /db_xref="taxon:7227"
                    /clone="BACR19D16"
                    /clone_lib="RPCI-98"
                    /note="end : TET3"
BASE COUNT      120 a     61 c     61 g    172 t    511 others
ORIGIN

  Query Match             14.7%;  Score 64.4;  DB 17;  Length 925;
  Best Local Similarity   14.2%;  Pred. No. 0.00013;
  Matches   50;  Conservative  168;  Mismatches  134;  Indels    0;  Gaps    0;

Qy         41 CAGCAGCCGGACTCGACCCCACCGAACACCCCCTACTCGGCGCCACATTGGAACTGGCGA 100
              | :: :|::: | |  |||| :|:  | ::::    :: :::| |   :: :|  :|:
Db        574 CCSSCSCSSSSCSCCBCCCCCSCSSYCCSSSBSSSKCSSTSBSCSCCCSSKSVCGTSCSS 633

Qy        101 CTGACGGTGGAGCGCTTCTTGCAGGGCGCTTGTCTTTGAGGTCGCATCCGTGGCTGGCTG 160
              : : ::: :: ::::| : |:  ::: ::  : : :|  ::|: :   :   |  :|:||
Db        634 SSSCSSSSSSTSSSSTSSSTSKSSSSGSSSSSSSSYTTSKSTSASGSGSWSAGGGSGSTG 693

Qy        161 ACCATGCCGTCGGCGGCACGGTGCTGCTGTCGGGCGCCACCTTCCTCGAACTCGCCCTTC 220
                :  ::::  : :::::::| ::  : : :::::|:: :: |:::::  :|:::::  :
Db        694 STSSSSSSSSTSTSSSSVSSGSKSSTBSSGSBSSSGSSSSSSSTSSBBSCTSTSSSSSSSY 753

Qy        221 ATGCGGGCACATACGTGGGCTGCGACCGAGTGGATGAGCTGACGCTGCATGCGCCGCTGG 280
              :  : | |  ::::::  :::  :  :|:|  : ::   :::::  | |:|| :
Db        754 SSSTCSCCTCCCSYSYSSSTSSSSSTSWGSTSGSSSSSVGTSSSSDSTSTCCSCCCYMCT 813

Qy        281 TGGTTCCTGTGGATGGGGGTGTGAGTGTGCAGGTTGGGGTTGCGGCTGCGGATGGGGAGG 340
              :|:::: : :   ||:: : | | :|:|  |  |:::| |::|  :  :  :::  ::
Db        814 CCSTYBMBCYTSTSCGGSSSSSGKGGVTKCGCGGCGSSSTNGMBGTSSACSSSSSCSSS 873

Qy        341 GGCGGCGTTTGGTGAGTGTGTATGCGCGGGGTGGGAGTGCTTGTGGTGGGGG 392
              :::::::   :: ::: : :   :: ::: ::::: | ::  : :| |:::|
Db        874 SVSSSSKSSASSSSVSSSGSSGVSSNSSSASKSSSSGSVSSGSGSGSGSVSG 925


RESULT 5
CNS016TH/c
LOCUS       CNS016TH                 1101 bp    DNA     linear   GSS 26-JUL-1999
DEFINITION  Drosophila melanogaster genome survey sequence SP6 end of BAC
            BACN16L02 of DrosBAC library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL107183
VERSION     AL107183.1  GI:5626028
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1101)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the European Drosophila Genome Project (EDGP) -
            http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
            library (Dros BAC) was made by Alain Billaud at CEPH (Centre
            d'Etude du Polymorphisme Humain) with funding provided by a MRC
            project grant. The DNA was prepared from embryos by Alain Bucheton
            and Genevieve Payan. It has been constructed in the vector
            pBeloBAC11.
FEATURES            Location/Qualifiers
     source         1. .1101
                    /organism="Drosophila melanogaster"
                    /db_xref="taxon:7227"
                    /clone="BACN16L02"
                    /clone_lib="DrosBAC"
                    /plasmid="pBeloBAC11"
                    /note="end : SP6"
BASE COUNT      196 a    304 c    161 g    149 t    291 others
```

```
ORIGIN

  Query Match            14.7%;  Score 64.2;  DB 17;  Length 1101;
  Best Local Similarity  46.7%;  Pred. No. 0.00015;
  Matches   93;  Conservative   29;  Mismatches   77;  Indels    0;  Gaps    0;

Qy        222 TGCGGGCACATACGTGGGCTGCGACCGAGTGGATGAGCTGACGCTGCATGCGCCGCTGGT 281
              |||||:  | |  || ||        :: : ||  : | :|  || |  |  || |  ||
Db        332 TGCGGKGGCGTTGGTTGGTGTGTGGSKKRGGGTAKTGTKGGGGCGGTKTCTGCKGGGGGG 273

Qy        282 GGTTCCTGTGGATGGGGGTGTGAGTGTGCAGGTTGGGGTTGCGGCTGCGGATGGGGAGGG 341
              |||    | ||  ||||| : | | |||  |: :|::|: | ||  | |  ||||| ||:
Db        272 GGTGGGGGGGGGGGGGGGKGGGGGGTGGGGKGKGKKGKGGGGGGCGGGTKTGGGGGGGK 213

Qy        342 GCGGCGTTTGGTGAGTGTGTATGCGCGGGGTGGGAGTGCTTGTGGTGGGGGTGGTGCGTC 401
              : || ::::|||| | | |||:| | |||: :|. : | : | |  ||||| :|:| |
Db        212 KGGGGKKKKGGTGKGGGGGTAKGGGGGGGKGKGTTKGGGKGGGGTGGGGGGGKGKGGGGG 153

Qy        402 GGGTGGGGTGTGGACGTGT 420
              .|||:::|| | ||  | |:
Db        152 GGGKKKGGGGGGGGGGGK 134


RESULT 6
AG076826/c
LOCUS       AG076826               796 bp    DNA     linear   GSS 03-NOV-2001
DEFINITION  Pan troglodytes DNA, clone: PTB-071C09.R, genomic survey sequence.
ACCESSION   AG076826
VERSION     AG076826.1  GI:16628628
KEYWORDS    GSS.
SOURCE      Pan troglodytes male lymphoblast DNA, clone_lib:PTB Chimpanzee Male
            BAC Library clone:PTB-071C09.R.
  ORGANISM  Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE   1
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     BAC end sequences of Library PTB
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 796)
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            (E-mail:chimpbes@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT     Clones are derived from the chimpanzee BAC library PTB This BAC end
            was generated during the R&D process and may have higher chance of
            clone tracking errors.
             PRIMERS
                 Sequencing: M13Rev
             LIBRARY
                 Vector    : pKS145
                 R.Site 1  : SacI
                 R.Site 2  : SacI.
FEATURES             Location/Qualifiers
     source          1. .796
                     /organism="Pan troglodytes"
                     /db_xref="taxon:9598"
                     /clone="PTB-071C09.R"
                     /sex="male"
                     /cell_type="lymphoblast"
                     /clone_lib="PTB Chimpanzee Male BAC Library"
BASE COUNT      125 a    444 c    156 g     36 t     35 others
ORIGIN

  Query Match            14.6%;  Score 64;  DB 17;  Length 796;
  Best Local Similarity  51.8%;  Pred. No. 0.00016;
  Matches  175;  Conservative    0;  Mismatches  155;  Indels    8;  Gaps    1;

Qy         95 TGGCGACTGACGGTGGAGCGCTTCTTGCAGGGCGCTTGTCTTTGAGGTCGCATCCGTGGC 154
              ||||| | | |||||| ||||  |  |  |||  |    |          ||  |  | | ||
Db        750 TGGCGGCCGGCGGTGGCGCGCGGCGGGGGGGGGGGTGGGGGGGGTGGGCGGGGGCGGGGGG 691

Qy        155 TGGCTGACCATGCCGTCGGCGGCACGGTGCTGCTGTCGGGCGCCACCTTCCTCGAACTCG 214
              ||  |     || | ||||||   || |  |  | || |||    |      |       |
Db        690 GGGGGGGGTTGGCGGGCGGCGGGGGGGCGTGGGGGGGGGCCGCTTGGTGGGGGGGTGGGG 631

Qy        215 CCCTTCATGCGGGCACATACGTGGGCTGCGACCGAGTGGATGAGCTGACGCTGCATGCGC 274
              |   || ||          |  | | |  | ||||  | | ||  |  |   ||||
Db        630 GGGTGTGTGGGGCGGGGGGGGGCCGGGGGGGGGGGGTGGGGGGGGTGGGGGCGTTGGCGC 571

Qy        275 CGCTGGTGGTTCCTGTGGATGGGGGTGTGAGTGTGCAGGTTGGGGTTGCGGCTGCGGATG 334
              ||  || ||||         |||| ||| | | |||  || |||||| |||| | || |
Db        570 CGGGGGGGGTT--------TGGGCGTGGGGGGGTGGTGGGTGGGGTCGCGGGGGGGGGGGG 519

Qy        335 GGGAGGGGCGGCGTTTGGTGAGTGTGTATGCGCGGGGTGGGAGTGCTTGTGGTGGGGGTG 394
              ||| | || |||||  || | | | |   | |||| || | |   |||| |||| |
Db        518 GGGGGCGGGGGCGTGGGGGGGGGGCGGGCGGGGGGGGCCGGGGGGGGGGTGGGCGGGGCG 459

Qy        395 GTGCGTCGGGTGGGGTGTGGACGTGTCATGCCTCGGGG 432
              ||| |  |||| ||| | ||  | |    ||  |||||
Db        458 GTGGGGGGGGTTGGGGGGGGGGGGGGGGGGGCGGCGGGG 421


RESULT 7
BQ068225
LOCUS       BQ068225               985 bp    mRNA    linear   EST 02-APR-2002
DEFINITION  AGENCOURT_6794375 NIH_MGC_121 Homo sapiens cDNA clone IMAGE:5770460
            5', mRNA sequence.
ACCESSION   BQ068225
VERSION     BQ068225.1  GI:19897271
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 985)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Life Technologies, Inc.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM12833  row: i  column: 21
            High quality sequence stop: 679.
FEATURES             Location/Qualifiers
     source          1. .985
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5770460"
                     /clone_lib="NIH_MGC_121"
                     /lab_host="DH10B"
                     /note="Organ: brain; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: EcoRV (destroyed); RNA source anonymous pool of 3
                     fetal brains, female age 20 weeks, female age 24 weeks,
                     and male age 26 weeks.  Library is oligo-dT primed and
                     directionally cloned (EcoRV site is destroyed upon
                     cloning).  Average insert size 1.7 kb, insert size range
                     0.7-3.5 kb.  Library is normalized and enriched for
                     full-length clones and  was constructed by C. Gruber
                     (Invitrogen).  Research Genetics  tracking code 017. Note:
                     this is a NIH_MGC Library."
BASE COUNT      111 a    288 c    355 g    139 t     92 others
ORIGIN

  Query Match            14.6%;  Score 64;  DB 14;  Length 985;
```

```
  Best Local Similarity   39.5%;  Pred. No. 0.00017;
  Matches  109;  Conservative    0;  Mismatches  167;  Indels    0;  Gaps    0;

Qy        160 GACCATGCCGTCGGCGGCACGGTGCTGCTGTCGGGCGCCACCTTCCTCGAACTCGCCCTT 219
              || || |    ||| ||  | |||  | |||  |||
Db        710 GAGCAGGGGCCCGGTGGGTCCGTGGGGGTGTGGGGGNNNNNNNNNNNNNNNNNNNNNNN 769

Qy        220 CATGCGGGCACATACGTGGGCTGCGACCGAGTGGATGAGCTGACGCTGCATGCGCCGCTG 279
                          | | |        || | | | | | | | | | |  |
Db        770 NNNNNNNNNNNNNGNNNNGNGGGGNNNNNNGGGGGGNGNNGGGGGGGGGNGGGNNGNNG 829

Qy        280 GTGGTTCCTGTGGATGGGGGTGTGAGTGTGCAGGTTGGGGTTGCGGCTGCGGATGGGGAG 339
              | |    | ||  ||||| | | |   |   | ||||  | ||  | ||   ||
Db        830 GGNGNNNNGGGGGGGGGGGGGGGGGGNGGGGNGGGGGGGGGGGGGNGGNGGGNNGGNNN 889

Qy        340 GGGCGGCGTTTGGTGAGTGTGTATGCGCGGGGTGGGAGTGCTTGTGGTGGGGGTGGTGCG 399
              ||| ||     || | | | |   | | ||| ||| |     |    ||||| || | |
Db        890 GGGGGGGNNGGGGGGGGNGGGGGGGGGGGGGNGGGGGGGNNNNGNNNNGGGGGGGGGGGG 949

Qy        400 TCGGGTGGGGTGTGGACGTGTCATGCCTCGGGGGTG 435
                |||  |||| |  |    |    |    ||||| |
Db        950 NNGGGGGGGGGGNNGNNNGGGGGGGNGGGGGGGGG 985


RESULT 8
CNS0073W
LOCUS       CNS0073W                922 bp    DNA     linear   GSS 03-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence TET3 end of BAC #
            BACR14D09 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL066784
VERSION     AL066784.1  GI:4945247
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 922)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1. .922
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone="BACR14D09"
                     /clone_lib="RPCI-98"
                     /note="end : TET3"
BASE COUNT      223 a     95 c    109 g    221 t    274 others
ORIGIN

  Query Match             14.4%;  Score 63;  DB 17;  Length 922;
  Best Local Similarity   19.3%;  Pred. No. 0.00027;
  Matches   67;  Conservative  137;  Mismatches  144;  Indels    0;  Gaps    0;

Qy         88 TTGGAACTGGCGACTGACGGTGGAGCGCTTCTTGCAGGGCGCTTGTCTTTGAGGTCGCAT 147
              :|::   |: |  ::    ::: | |  | : :  :|  : |||| ::|  :|| :  :
Db        565 KTKKTTKTSTGGGKKKKTTYKBKTGTGTSTGKGKKGKGTKKTTTGTGKKTTGKGTKKTTK 624

Qy        148 CCGTGGCTGGCTGACCATGCCGTCGGCGGCACGGTGCTGCTGTCGGGCGCCACCTTCCTC 207
                :::| ::: |:    ::      |:  |   ||:: |: |:: ::  :      |
Db        625 KTKKKGTKKKKTKTTKGKKKGTGKGKKTGTTGGGKKKTKKTKKKKKTGKKKKKKTGKKGK 684

Qy        208 GAACTCGCCCTTCATGCGGGCACATACGTGGGCTGCGACCGAGTGGATGAGCTGACGCTG 267
              :   : :   |: :| ::    |  :| :: |: :   :  : : || : ::  : |:
Db        685 KGTKKTKTKTTKGGKGGKKTKKKKTTKKTTKKTTKTKKKKKTTKTKKTGTKKKKTTKTTK 744

Qy        268 CATGCGCCGCTGGTGGTTCCTGTGGATGGGGGTGTGAGTGTGCAGGTTGGGGTTGCGGCT 327
              :| :  :  :: |:|:    ::: |||:|::|:: :::::  :|:::: |:    : :
Db        745 KGKGTKKGKKGKKGGKTKGTGTKKKGTGGKGKKGKKGKKKKKKKKGKKKKTGKGTGTKKK 804

Qy        328 GCGGATGGGGAGGGGCGGCGTTTGGTGAGTGTGTATGCGCGGGGTGGGAGTGCTTGTGGT 387
              |  : ::  :::: :: :|:::||: | :|:: |: | :::||::: :|  :|||| :
Db        805 TKGTGKTKKTKKKKKKKKTKTKKKGTKKGGKTKKGTKKGKKKKGTKKKNKTTKKTGTGTK 864

Qy        388 GGGGGTGGTGCGTCGGGTGGGGTGTGGACGTGTCATGCCTCGGGGGTG 435
              :::::::|:| |: ::: :::|:::|:  : ::  ::  : |||||:|
Db        865 KKKKKKKGKGKGKKKKKGKKKGKKKGKGGKGKKKGKKGKKKGGGGGKG 912


RESULT 9
CNS008N8/c
LOCUS       CNS008N8                1094 bp    DNA     linear   GSS 03-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence T7 end of BAC #
            BACR17G15 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL052088
VERSION     AL052088.1  GI:4933940
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1094)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1. .1094
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone="BACR17G15"
                     /clone_lib="RPCI-98"
                     /note="end : T7"
BASE COUNT      319 a    407 c    105 g     80 t    183 others
ORIGIN

  Query Match             14.4%;  Score 63;  DB 17;  Length 1094;
  Best Local Similarity   57.6%;  Pred. No. 0.00028;
  Matches   87;  Conservative   12;  Mismatches   52;  Indels    0;  Gaps    0;
```

```
Qy        288 TGTGGATGGGGGTGTGAGTGTGCAGGTTGGGGTTGCGGCTGCGGATGGGGAGGGGCGGCG 347
              || |   |:| | ||| ||| |  ||| ||||::| ||  | || ||||| | || :| |
Db        388 TGGGTTGGKGTGGGTGGGTGGGGGGGTGGGGGKKGGGGGGGGGGGTGGGGGGTGGGKGGG 329

Qy        348 TTTGGTGAGTGTGTATGCGCGGGGTGGGAGTGCTTGTGGTGGGGGTGGTGCGTCGGGTGG 407
              |||| | ||| . | | |||:|||  | |  ||:|| ||||::|||: :  |||| |
Db        328 GGGGGTGGGGGTGGGGKGGGGGKTGGTGGGGGGTGKGGGGGGGKKGGTKKKGGGGGTTG 269

Qy        408 GGTGTGGACGTGTCATGCCTCGGGGGTGCTG 438
              || :||  | |    |  | ||||||| |:
Db        268 GGGGKGGGGGGGGGGGGGTTTGGGGGTGGTK 238


RESULT 10
BM468388/c
LOCUS       BM468388                1168 bp    mRNA    linear   EST 05-FEB-2002
DEFINITION  AGENCOURT_6475677 NIH_MGC_72 Homo sapiens cDNA clone IMAGE:5578074
            5', mRNA sequence.
ACCESSION   BM468388
VERSION     BM468388.1  GI:18517430
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1168)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC/DCTD/DTP
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM12332  row: i  column: 19
            High quality sequence stop: 291.
FEATURES             Location/Qualifiers
     source          1. .1168
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5578074"
                     /clone_lib="NIH_MGC_72"
                     /tissue_type="melanotic melanoma"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: skin; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally.  Primer: Oligo dT.
                     Average insert size 2 kb.   Library constructed by Life
                     Technologies."
BASE COUNT       72 a    611 c    310 g     48 t    127 others
ORIGIN

  Query Match             14.3%;  Score 62.8;  DB 13;  Length 1168;
  Best Local Similarity   44.5%;  Pred. No. 0.00032;
  Matches  190;  Conservative    0;  Mismatches  237;  Indels    0;  Gaps    0;

Qy          9 CGAAAGCACACAGCCCGGTGCCGGCAACGTGTCAGCAGCCGGACTCGACCCCACCGAACA 68
              ||   || | | |||||| ||| |  || | | |  ||| | | || |||| ||    |
Db        870 CGGCGGCGCGCCGCCCGGCCCCGCCCCCGGGGCCGGCGCCCGCCGCGGCCCCGCCCGCCC 811

Qy         69 CCCCCTACTCGGCGCCACATTGGAACTGGCGACTGACGGTGGAGCGCTTCTTGCAGGGCG 128
                |       | |||  |   |      |||   |  || |  |       |   || |
Db        810 GCNGGGCGGGGCCGCGGCGCGGCCGGCCGCGCGGGGGGGGGGNNGGNGGGGGGGGNNGGGG 751

Qy        129 CTTGTCTTTGAGGTCGCATCCGTGGCTGGCTGACCATGCCGTCGGCGGCACGGTGCTGCT 188
              | | |  | |  |    | ||| ||  |       ||  ||  |   || |
Db        750 CGGGCCGGGGGGNNGGGGGGGGGGCCGGGGGGNNGGNGCGGGGGNNGGGGGGGGNNNGG 691

Qy        189 GTCGGGCGCCACCTTCCTCGAACTCGCCCTTCATGCGGGCACATACGTGGGCTGCGACCG 248
```

```
              | ||||          |     |      | |||      | |||  | |   |
Db        690 GGCGGGGNGGGGNGGNNNGGGGGGGGNGGGGGGGGGGGGGGGGGGGGGGGGGNGG 631

Qy        249 AGTGGATGAGCTGACGCTGCATGCGCCGCTGGTGGTTCCTGTGGATGGGGGTGTGAGTGT 308
              | ||  | | |  |  | |  | |  || ||      | ||  ||||| | | | |
Db        630 GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 571

Qy        309 GCAGGTTGGGGTTGCGGCTGCGGATGGGGAGGGGCGGCGTTTGGTGAGTGTGTATGCGCG 368
              |  ||  ||||  | ||  | ||  |||| |||| || |   || | | | |   | | |
Db        570 GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 511

Qy        369 GGGTGGGAGTGCTTGTGGTGGGGGTGGTGCGTCGGGTGGGGTGTGGACGTGTCATGCCTC 428
              ||| ||| |     | || ||||| || | |  ||| || | | ||  | |     |
Db        510 GGGGGGGGGGNGGGGGGGGGGGGGGGGGGGGGGGGGGGGGNGGGGGGGGGGGGGGGGGG 451

Qy        429 GGGGGTG 435
              ||||| |
Db        450 GGGGGGG 444


RESULT 11
AZ191837
LOCUS       AZ191837                797 bp    DNA     linear   GSS 30-AUG-2000
DEFINITION  SP_1020_A2_A08_T7A Strongylocentrotus purpuratus, purple sea urchin
            , sperm genomic BAC library Strongylocentrotus purpuratus genomic
            clone Plate=1020 Col=16 Row=A, DNA sequence.
ACCESSION   AZ191837
VERSION     AZ191837.1  GI:8375016
KEYWORDS    GSS.
SOURCE      Strongylocentrotus purpuratus.
  ORGANISM  Strongylocentrotus purpuratus
            Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
            Echinoidea; Euechinoidea; Echinacea; Echinoida;
            Strongylocentrotidae; Strongylocentrotus.
REFERENCE   1  (bases 1 to 797)
  AUTHORS   Cameron,R.A., Mahairas,G., Rast,J.P., Martinez,P., Biondi,T.R.,
            Swartzell,S., Wallace,J.C., Poustka,A.J., Livingston,B.T., Wray
            ,G.A., Ettensohn,C.A., Lehrach,H., Britten,R.J, Davidson,E.H. and
            Hood,L.
  TITLE     A sea urchin genome project: Sequence scan, virtual map, and
            additional resources
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (17), 9514-9518 (2000)
  MEDLINE   20402566
COMMENT     Contact: Cameron, RA, Davidson, EH, Hood, L
            Division of Biology 156-29
            California Institute of Technology
            Pasadena California 91125, USA
            Tel: (626) 395-8421
            Fax: (626) 793-3047
            Email: acameron@caltech.edu
            Plate: 1020  row: A  column: 16
            Seq primer: T7
            Class: BAC ends
            High quality sequence stop: 797.
FEATURES             Location/Qualifiers
     source          1. .797
                     /organism="Strongylocentrotus purpuratus"
                     /db_xref="taxon:7668"
                     /clone="Plate=1020 Col=16 Row=A"
                     /clone_lib="Strongylocentrotus purpuratus, purple sea
                     urchin, sperm genomic BAC library"
                     /note="Organ: sperm; Vector: BACe3.6; BAC Clones in E-Coli
                     DH10B"
BASE COUNT       98 a    124 c    312 g    161 t    102 others
ORIGIN

  Query Match             14.3%;  Score 62.6;  DB 17;  Length 797;
  Best Local Similarity   38.5%;  Pred. No. 0.00033;
  Matches  107;  Conservative    0;  Mismatches  171;  Indels    0;  Gaps    0;

Qy        158 CTGACCATGCCGTCGGCGGCACGGTGCTGCTGTCGGGCGCCACCTTCCTCGAACTCGCCC 217
              | ||||  |  || |||||  | || |  |
Db        470 CCGACCCCGGAGTTGGCGGGAGGGGGGGGNNNNNNNNNNNNNNNNNNNNNNNNNNNN 529
```

```
Qy        218 TTCATGCGGGCACATACGTGGGCTGCGACCGAGTGGATGAGCTGACGCTGCATGCGCCGC 277
Db        530 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNGNNNNNNNNNNNNNNGNNGNNNGGGNGGG 589

Qy        278 TGGTGGTTCCTGTGGATGGGGGTGTGAGTGTGCAGGTTGGGGTTGCGGCTGCGGATGGGG 337
Db        590 GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGNGGGGGGGGGNGGNGGGGGGNGGGGGGG 649

Qy        338 AGGGGCGGCGTTTGGTGAGTGTGTATGCGCGGGGTGGGAGTGCTTGTGGTGGGGGTGGTG 397
Db        650 GGGGGGGNGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGNNGGGGGGGGGGGGGGG 709

Qy        398 CGTCGGGTGGGGTGTGGACGTGTCATGCCTCGGGGGTG 435
Db        710 GGGGGGGGGGNGNGGGGNNNGGGGGGGGNNNGGGGGGG 747


RESULT 12
CNS006U0
LOCUS       CNS006U0                  884 bp    DNA     linear   GSS 03-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence T7 end of BAC #
            BACR14N21 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL065923
VERSION     AL065923.1  GI:4944891
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 884)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1. .884
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone="BACR14N21"
                     /clone_lib="RPCI-98"
                     /note="end : T7"
BASE COUNT      230 a     62 c    139 g    124 t    329 others
ORIGIN

  Query Match             14.2%;  Score 62.4;  DB 17;  Length 884;
  Best Local Similarity   15.2%;  Pred. No. 0.00037;
  Matches   31;  Conservative  113;  Mismatches   60;  Indels    0;  Gaps    0;

Qy        235 GTGGGCTGCGACCGAGTGGATGAGCTGACGCTGCATGCGCCGCTGGTGGTTCCTGTGGAT 294
Db        656 GGGGGKKKGGTKTGGKDGTGGKKKTTKKKKKKKKGTKGKGKGKTKKKSGTKKKKGGTGKG 715

Qy        295 GGGGGTGTGAGTGTGCAGGTTGGGGTTGCGGCTGCGGATGGGGAGGGGCGGCGTTTGGTG 354
Db        716 TGTKGTGTKKKKGGKGKGGKKKGGKGKKGKKKKKGKKKKKKKKKGKKKKKTGKKTKGKKG 775

Qy        355 AGTGTGTATGCGCGGGGTGGGAGTGCTTGTGGTGGGGGTGGTGCGTCGGGTGGGGTGTGG 414
Db        776 GTKKKGKKKGKKTKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKTKKKKKKKKKK 835

Qy        415 ACGTGTCATGCCTCGGGGGTGCTG 438
Db        836 CKKKKKKKKKKKKKKKKKKKKKK 859


RESULT 13
BQ433302/c
LOCUS       BQ433302                  886 bp    mRNA    linear   EST 24-MAY-2002
DEFINITION  AGENCOURT_7723727 NIH_MGC_72 Homo sapiens cDNA clone IMAGE:6050702
            5', mRNA sequence.
ACCESSION   BQ433302
VERSION     BQ433302.1  GI:21172378
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 886)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC/DCTD/DTP
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM13303  row: f  column: 15
            High quality sequence stop: 326.
FEATURES             Location/Qualifiers
     source          1. .886
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:6050702"
                     /clone_lib="NIH_MGC_72"
                     /tissue_type="melanotic melanoma"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: skin; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally.  Primer: Oligo dT.
                     Average insert size 2 kb.   Library constructed by Life
                     Technologies."
BASE COUNT       75 a    520 c    135 g    116 t     40 others
ORIGIN

  Query Match             14.2%;  Score 62;  DB 14;  Length 886;
  Best Local Similarity   47.8%;  Pred. No. 0.00045;
  Matches  131;  Conservative    0;  Mismatches  143;  Indels    0;  Gaps    0;

Qy        162 CCATGCCGTCGGCGGCACGGTGCTGCTGTCGGGCGCCACCTTCCTCGAACTCGCCCTTCA 221
Db        762 CCGGGGCGGCGCGGGCACGCGGCGCGNCCGGCGCGGCGCCCCGCCCGGGCGCCGGCGGCG 703

Qy        222 TGCGGGCACATACGTGGGCTGCGACCGAGTGGATGAGCTGACGCTGCATGCGCCGCTGGT 281
Db        702 CNGCGNCNCGGGGCGGGGCGGCGGCCGCGGCCNGGGGGNGGGGGGNNGGGGGGGGGGGGG 643

Qy        282 GGTTCCTGTGGATGGGGGTGTGAGTGTGCAGGTTGGGGTTGCGGCTGCGGATGGGGAGGG 341
Db        642 GGGGNNNGGGGGGGGGGGGGGGGNGGNNGGGGNGGGGCGGGGGGGGGGGGGGGGNGNG 583

Qy        342 GCGGCGTTTGGTGAGTGTGTATGCGCGGGGTGGGAGTGCTTGTGGTGGGGGTGGTGCGTC 401
Db        582 GGGGGGGGGGGGGGGGGGGGGGGGGNNNGGGNGGGGGGGGGGGGGGGGGGGGGGGGG 523
```

```
Qy        402 GGGTGGGGTGTGGACGTGTCATGCCTCGGGGGTG 435
              ||| |||| | | | | |     |  |||||  |
Db        522 GGGGGGGGGGGNGNGGGGGGGGNNNNGGGGGGG 489


RESULT 14
CNS00397/c
LOCUS       CNS00397                1101 bp    DNA     linear   GSS 03-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence T7 end of BAC #
            BACR08A10 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL063912
VERSION     AL063912.1  GI:4941769
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1101)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES              Location/Qualifiers
     source           1. .1101
                      /organism="Drosophila melanogaster"
                      /db_xref="taxon:7227"
                      /clone="BACR08A10"
                      /clone_lib="RPCI-98"
                      /note="end : T7"
BASE COUNT      111 a    286 c    221 g    241 t    242 others
ORIGIN

  Query Match             14.1%;  Score 61.6;  DB 17;  Length 1101;
  Best Local Similarity   32.9%;  Pred. No. 0.00059;
  Matches  117;  Conservative   78;  Mismatches  161;  Indels    0;  Gaps    0;

Qy         80 GCGCCACATTGGAACTGGCGACTGACGGTGGAGCGCTTCTTGCAGGGCGCTTGTCTTTGA 139
              ||||     :::  :: |||  ::  :|::| | : :: ::|  |:|:|||:|| ::::
Db       1099 GCGCKCGYSYSKCKYKCGCGYGKKBKBGKKGKGGKKYKGKKGTGGSGBGCTYGTTYBKKC 1040

Qy        140 GGTCGCATCCGTGGCTGGCTGACCATGCCGTCGGCGGCACGGTGCTGCTGTCGGGCGCCA 199
              |  |   : :| : | ::::|     :|::: :|:| :  :: ::  |:::|   | :
Db       1039 GTCCCKTYGYGCSTCGKBBYGGTGCSKCYKBGKGYGYSKTSKSKBCTCYSYCCCCCCYKC 980

Qy        200 CCTTCCTCGAACTCGCCCTTCATGCGGGCACATACGTGGGCTGCGACCGAGTGGATGAGC 259
              ||  :| |   | |:| |  | : :  ::   : : :| |  ::  |   : :
Db        979 CCCCBCCCCCCCCCSCGCCSCCBCSCYKSCTCYCYCYGYGTCKYYTCKTYKCBYKCTTYK 920

Qy        260 TGACGCTGCATGCGCCGCTGGTGGTTCCTGTGGATGGGGGTGTGAGTGTGCAGGTTGGGG 319
              ::  |. ::  |||::| :| :|||     |  |  ::|| ||| |:      |   | |
Db        919 BKBTGTCKSKSGCGBSGGKGYKGGTGTGGGGCGYGBSRGGCGTGKGKCGCGCGTCGWGCG 860

Qy        320 TTGCGGCTGCGGATGGGGAGGGGCGGCGTTTGGTGAGTGTGTATGCGCGGGGTGGGAGTG 379
              : |  :  | ||  ||:| |||: || |   || | : | |   | | | || |:| | |
Db        859 WSGGARGGGAGGGGGGRGGGGGSGGGGGGGSGGGGSGGGGRGGGGGGGCGGSGSGGGGG 800
```

```
Qy        380 CTTGTGGTGGGGGTGGTGCGTCGGGTGGGGTGTGGACGTGTCATGCCTCGGGGGTG 435
                | :| ||||| || |::  ||| |||| | |||  | |     |    ||||| |
Db        799 GGGGGKGGGGGGGSGGGGVSGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 744


RESULT 15
BQ653252/c
LOCUS       BQ653252                1062 bp    mRNA    linear   EST 15-JUL-2002
DEFINITION  AGENCOURT_8341660 NIH_MGC_100 Homo sapiens cDNA clone IMAGE:6267901
            5', mRNA sequence.
ACCESSION   BQ653252
VERSION     BQ653252.1  GI:21777424
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1062)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: CGAP (Stanford)
             cDNA Library Preparation: Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM2438  row: p  column: 14
            High quality sequence stop: 431.
FEATURES              Location/Qualifiers
     source           1. .1062
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /clone="IMAGE:6267901"
                      /clone_lib="NIH_MGC_100"
                      /tissue_type="hepatocellular carcinoma, cell line"
                      /lab_host="DH10B (phage-resistant)"
                      /note="Organ: liver; Vector: pOTB7; Site_1: XhoI; Site_2:
                      EcoRI; cDNA made by oligo-dT priming. Directionally cloned
                      into EcoRI/XhoI sites using the following 5' adaptor:
                      GGCACGAG(G). Size-selected >500bp for average insert size
                      1.8kb. Library constructed by Ling Hong in the laboratory
                      of Gerald M. Rubin (University of California, Berkeley)
                      using ZAP-cDNA synthesis kit (Stratagene) and Superscript
                      II RT (Life Technologies). Note: this is a NIH_MGC
                      Library."
BASE COUNT      185 a    465 c    201 g    129 t     82 others
ORIGIN

  Query Match             14.0%;  Score 61.4;  DB 14;  Length 1062;
  Best Local Similarity   48.3%;  Pred. No. 0.00065;
  Matches  143;  Conservative    0;  Mismatches  153;  Indels    0;  Gaps    0;

Qy        140 GGTCGCATCCGTGGCTGGCTGACCATGCCGTCGGCGGCACGGTGCTGCTGTCGGGCGCCA 199
              || |||||| | || | | || ||  || | | |  ||  || | |  | | ||| || |
Db        836 GGGCGCATCGGGGGGTCGGTGTCCCGGCAGCCTGGCGCCNGGGGGCGGGGCCGGCCGGCG 777

Qy        200 CCTTCCTCGAACTCGCCCTTCATGCGGGCACATACGTGGGCTGCGACCGAGTGGATGAGC 259
              |       |    ||||     ||| |      ||  |   |||   | | ||  |
Db        776 GCGGGGCGGGCTGGGCCCCGGGAGCGCGGCNGGGCGGCGCGCGCGGGGGNGGGGNGGGCG 717

Qy        260 TGACGCTGCATGCGCCGCTGGTGGTTCCTGTGGATGGGGGTGTGAGTGTGCAGGTTGGGG 319
              |  |  |   |    |  || ||     | || |||| | | | |   |  ||   |||
Db        716 GGGGGGGGGGGGGNNGGGGGGGGGGGNGGGNGGGGGTGGGTGGGCGGGGCGGGGGGGGNGGG 657

Qy        320 TTGCGGCTGCGGATGGGGAGGGGCGGCGTTTGGTGAGTGTGTATGCGCGGGGTGGGAGTG 379
                |  ||  |||| |||| || |   ||   | | |   | | |||| ||| | |
Db        656 GGNGNGGGGGGGNAGGGGGGGGGGGGGGGGGNGGGNGGGGGGGGGGGGGGGGGGNGNG 597
```

```
Qy        380 CTTGTGGTGGGGGTGGTGCGTCGGGTGGGGTGTGGACGTGTCATGCCTCGGGGGTG 435
              | || ||||| || | |  ||| |||| | ||  | |    |    ||| | |
Db        596 GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGNGGG 541
```

Search completed: June 18, 2003, 01:00:29
Job time : 564.563 secs

THIS PAGE BLANK (USPTO)

```
Qy      1861 ANYAAANAYLDALAHHRHTHHLPATSIAWGTWQGNGLADSDKARAYLDRRGFRPMSPELA 1920
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1861 ANYAAANAYLDALAHHRHTHHLPATSIAWGTWQGNGLADSDKARAYLDRRGFRPMSPELA 1920

Qy      1921 TAAVTQAIADTERPYVVIADIDWSKIEHTSQTSDLVSAARAREPAVQRPTPPAELHKTLA 1980
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1921 TAAVTQAIADTERPYVVIADIDWSKIEHTSQTSDLVSAARAREPAVQRPTPPAELHKTLA 1980

Qy      1981 HQTSADQRAALLELVRDHVAAVLRHADPKAIAPDQSFRALGFDSLTAVEFRNLLIKATGL 2040
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1981 HQTSADQRAALLELVRDHVAAVLRHADPKAIAPDQSFRALGFDSLTAVEFRNLLIKATGL 2040

Qy      2041 RLPVSLVFDHPTPAKLAVHLQNQLRGTAAESAPSAAAVTAEASVTEPIAIVGMACRFPGG 2100
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2041 RLPVSLVFDHPTPAKLAVHLQNQLRGTAAESAPSAAAVTAEASVTEPIAIVGMACRFPGG 2100

Qy      2101 VTSADDFWDLISSEQDAIGGFPTDRGWDLDTLYDPDPDHPGTCYTRNGGFLYDAGHFDAE 2160
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2101 VTSADDFWDLISSEQDAIGGFPTDRGWDLDTLYDPDPDHPGTCYTRNGGFLYDAGHFDAE 2160

Qy      2161 FFGISPREALAMDPQQRLLLETAWETIEHAGINPHTLHGTPTGVFTGTNGQDYALRVHNA 2220
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2161 FFGISPREALAMDPQQRLLLETAWETIEHAGINPHTLHGTPTGVFTGTNGQDYALRVHNA 2220

Qy      2221 GQSTDGFALTGTAGSVISGRISYTFGFEGPAVSVDTACSSSLVALHLACQALRAGECSMA 2280
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2221 GQSTDGFALTGTAGSVISGRISYTFGFEGPAVSVDTACSSSLVALHLACQALRAGECSMA 2280

Qy      2281 LAGGVTVMSSPGAFVEFSRQRGLAADGHCKAFSAAADGTGWGEGVGMLLVERLSDAHRNG 2340
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2281 LAGGVTVMSSPGAFVEFSRQRGLAADGHCKAFSAAADGTGWGEGVGMLLVERLSDAHRNG 2340

Qy      2341 HRVLAVVRGSAVNQDGASNGLTAPNGPSQQRVIRQALANAGLSAGDVDAVEAHGTGTTLG 2400
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2341 HRVLAVVRGSAVNQDGASNGLTAPNGPSQQRVIRQALANAGLSAGDVDAVEAHGTGTTLG 2400

Qy      2401 DPIEAQALLATYGQDRAGEGPLWLGSVKSNVGHTQAAAGVAGVIKMVMALRHGLLPRTLH 2460
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2401 DPIEAQALLATYGQDRAGEGPLWLGSVKSNVGHTQAAAGVAGVIKMVMALRHGLLPRTLH 2460

Qy      2461 VDEPSPHVDWSAGAVQLLTETVPWPGGEGRLRRAGVSSFGVSGTNAHVILEEAPADDVPG 2520
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2461 VDEPSPHVDWSAGAVQLLTETVPWPGGEGRLRRAGVSSFGVSGTNAHVILEEAPADDVPG 2520

Qy      2521 GPPAGEGDAGSDDEAAAGSPGVWPWLVSAKSQPALRAQAQALHAHLTDHPGLDLADVGYT 2580
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2521 GPPAGEGDAGSDDEAAAGSPGVWPWLVSAKSQPALRAQAQALHAHLTDHPGLDLADVGYT 2580

Qy      2581 LAHARAVFDHRATLIAADRDTFLQALQALAAGEPHPAVIHSSAPGGTGTGEAAGKTAFIC 2640
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2581 LAHARAVFDHRATLIAADRDTFLQALQALAAGEPHPAVIHSSAPGGTGTGEAAGKTAFIC 2640

Qy      2641 SGQGTQRPGMAHGLYHTHPVFAAALNDICTHLDPHLDHPLLPLLTQNDNDNEDAAALLQQ 2700
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2641 SGQGTQRPGMAHGLYHTHPVFAAALNDICTHLDPHLDHPLLPLLTQNDNDNEDAAALLQQ 2700

Qy      2701 TRYAQPALFAFQVALHRLLTDGYHITPHYYAGHSLGEITAAHLAGILTLTDATTLITQRA 2760
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2701 TRYAQPALFAFQVALHRLLTDGYHITPHYYAGHSLGEITAAHLAGILTLTDATTLITQRA 2760

Qy      2761 TLMQTMPPGTMTTLHTTPHHITHHLTAHENDLAIAAINTPTSLVISGTPHTVQHITTLCQ 2820
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2761 TLMQTMPPGTMTTLHTTPHHITHHLTAHENDLAIAAINTPTSLVISGTPHTVQHITTLCQ 2820

Qy      2821 QQGIKTKTLPTNHAFHSPHTNPILNQLHQHTQTLTYHPPHTPLITANTPPDQLLTPHYWT 2880
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2821 QQGIKTKTLPTNHAFHSPHTNPILNQLHQHTQTLTYHPPHTPLITANTPPDQLLTPHYWT 2880

Qy      2881 QQARNTVDYATTTQTLHQHGVTTYIELGPDNTLTTLTHHNLPNPPTTTLTLTHPHHHPQT 2940
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2881 QQARNTVDYATTTQTLHQHGVTTYIELGPDNTLTTLTHHNLPNPPTTTLTLTHPHHHPQT 2940

Qy      2941 HLLTNLAKTTTTWHPHHYTHHDNQPHTHTHLDLPTYPFQHHHYWLESTQPGAGNVSAAGL 3000
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2941 HLLTNLAKTTTTWHPHHYTHHDNQPHTHTHLDLPTYPFQHHHYWLESTQPGAGNVSAAGL 3000

Qy      3001 DPTEHPLLGATLELATDGGALLAGRLSLRSHPWLADHAVGGTVLLSGATFLELALHAGTY 3060
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3001 DPTEHPLLGATLELATDGGALLAGRLSLRSHPWLADHAVGGTVLLSGATFLELALHAGTY 3060

Qy      3061 VGCDRVDELTLHAPLVVPVDGGVSVQVGVAAADGEGRRLVSVYARGGSACGGGGASGGVW 3120
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3061 VGCDRVDELTLHAPLVVPVDGGVSVQVGVAAADGEGRRLVSVYARGGSACGGGGASGGVW 3120

Qy      3121 TCHASGVLVEAAAGGVVVDGLAGVWPPRGAVAVDVDGVRDRLAGAGCVLGPVFSGLRAVW 3180
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3121 TCHASGVLVEAAAGGVVVDGLAGVWPPRGAVAVDVDGVRDRLAGAGCVLGPVFSGLRAVW 3180

Qy      3181 RDGGDLLAEVCLPEEAWGDAAGFGLHPALLDGVVQPLSVLLPGGTGFGEGAGFGEGVRVP 3240
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3181 RDGGDLLAEVCLPEEAWGDAAGFGLHPALLDGVVQPLSVLLPGGTGFGEGAGFGEGVRVP 3240

Qy      3241 AVWGGVSLHRAGVTGVRVRVSAVGRGGGREAVSVVVGDEAGVPVASVDRLELRPVDMGQL 3300
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3241 AVWGGVSLHRAGVTGVRVRVSAVGRGGGREAVSVVVGDEAGVPVASVDRLELRPVDMGQL 3300

Qy      3301 RAVSVSAGRRGSLYAVQWAEVGPVPVCGQAWAWHEDVGESGGGPVPGVVVLRCPDAGAGG 3360
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3301 RAVSVSAGRRGSLYAVQWAEVGPVPVCGQAWAWHEDVGESGGGPVPGVVVLRCPDAGAGG 3360

Qy      3361 GGGGGGGGGVGEVVGGVLGVVQGWLGLERFAGSRLVVVTRGAVVAGPEDGPVDVVGASVW 3420
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3361 GGGGGGGGGVGEVVGGVLGVVQGWLGLERFAGSRLVVVTRGAVVAGPEDGPVDVVGASVW 3420

Qy      3421 GLVRSAQAEHPDRFVLLDLDTDTGTDLDTGAGAGWGVDGGRVAAVVACGEPQLAVRGERL 3480
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3421 GLVRSAQAEHPDRFVLLDLDTDTGTDLDTGAGAGWGVDGGRVAAVVACGEPQLAVRGERL 3480

Qy      3481 LAARLKRLESSGDVPAQRSGDTRARRSDVPAQRSGGVPARRSVDVSGREVLPWLSGGSVL 3540
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3481 LAARLKRLESSGDVPAQRSGDTRARRSDVPAQRSGGVPARRSVDVSGREVLPWLSGGSVL 3540

Qy      3541 VTGGTGVLGAAVARHLAGVCGVRDLLLVSRRGPDAPGAEGLRAELAALGAEVRIVACDVG 3600
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3541 VTGGTGVLGAAVARHLAGVCGVRDLLLVSRRGPDAPGAEGLRAELAALGAEVRIVACDVG 3600

Qy      3601 ERREVVRLLEGVPAGCPLTGVVHAAGVLDDATIASLTPERLGTVFAAKVDAALLLDELTR 3660
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3601 ERREVVRLLEGVPAGCPLTGVVHAAGVLDDATIASLTPERLGTVFAAKVDAALLLDELTR 3660

Qy      3661 GMELSAFVLFSSAAGILGSAGQGNYAAANAALDALAYRRRAAGLPGVSLAWGLWEEASGM 3720
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3661 GMELSAFVLFSSAAGILGSAGQGNYAAANAALDALAYRRRAAGLPGVSLAWGLWEEASGM 3720

Qy      3721 TGHLAGTDHRRIIRSGLHPMSTPDALALFDAALALDRPVLLPADLRPAPPLPPLLQDLLP 3780
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3721 TGHLAGTDHRRIIRSGLHPMSTPDALALFDAALALDRPVLLPADLRPAPPLPPLLQDLLP 3780

Qy      3781 ATRRRTTRTTTTGGADNGAQLHARLAGQTHEQQHTTLLALVRSHIATVLGHTTPDTIPPD 3840
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3781 ATRRRTTRTTTTGGADNGAQLHARLAGQTHEQQHTTLLALVRSHIATVLGHTTPDTIPPD 3840

Qy      3841 RAFRDLGFDSLTAVELRNRLSRTTGLRLPTTLAFDHPNPTTLTHHLHTQLQPQPDNAVAP 3900
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3841 RAFRDLGFDSLTAVELRNRLSRTTGLRLPTTLAFDHPNPTTLTHHLHTQLQPQPDNAVAP 3900

Qy      3901 VLAELDKLESALSALDKTDSASERVTLRLKSLMLRWNAPQHPTAESADDDEKFTSATEAE 3960
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3901 VLAELDKLESALSALDKTDSASERVTLRLKSLMLRWNAPQHPTAESADDDEKFTSATEAE 3960

Qy      3961 IFKFIDNDLGLS 3972
             ||||||||||||
Db      3961 IFKFIDNDLGLS 3972

RESULT 2
```

AAG65264
ID AAG65264 standard; Protein; 3972 AA.
XX
AC AAG65264;
XX
DT 04-DEC-2001 (first entry)
XX
DE Streptomyces avermitilis protein SEQ ID NO: 4.
XX
KW Avermectin aglycone synthase; AAS; avermectin derivative;
KW drug production; vetinary drug; pesticide.
XX
OS Streptomyces avermitilis.
XX
PN WO200162939-A1.
XX
PD 30-AUG-2001.
XX
PF 23-FEB-2001; 2001WO-JP01381.
XX
PR 24-FEB-2000; 2000JP-0047405.
XX
PA (KYOW) KYOWA HAKKO KOGYO KK.
PA (KITA) KITASATO INST.
XX
PI Endo H, Yamaguchi H, Kanda Y, Hashimoto S, Omura S, Ikeda H;
XX
DR WPI; 2001-582053/65.
DR N-PSDB; AAH79277.
XX
PT New modified avermectin aglycone synthase derived from Streptomyces
PT avermectilis used in production of 22,23-dihydroavermectin B1a used in
PT drugs and pesticides -
XX
PS Claim 4; Page 167-180; 257pp; Japanese.
XX
CC The present invention relates to the production of modified derivatives
CC of avermectin aglycone synthase (AAS) derived from Streptomyces
CC avermitilis. The activity of an acyl carrier protein (ACP),
CC beta-ketoacyl carrier protein synthase (KS), acyltransferase (AT),
CC beta-ketoacyl carrier protein reductase (KR), dehydratase (DH), enoyl
CC reductase (ER) and/or thioesterase (TE) domain may be reduced or
CC suppressed. The process can be used in the production of drugs, vetinary
CC drugs and pesticides. The present sequence is an S. avermitilis protein.
XX
SQ Sequence 3972 AA;

Query Match 100.0%; Score 20829; DB 22; Length 3972;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 3972; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 VQRMDGGEEPRPAAGEVLGVADEADGGVVFVFPGQGPQWPGMGRELLDASDVFRESVRAC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 VQRMDGGEEPRPAAGEVLGVADEADGGVVFVFPGQGPQWPGMGRELLDASDVFRESVRAC 60

Qy         61 EAAFAPYVDWSVEQVLRDSPDAPGLDRVDVVQPTLFAVMISLAALWRSQGVEPCAVLGHS 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 EAAFAPYVDWSVEQVLRDSPDAPGLDRVDVVQPTLFAVMISLAALWRSQGVEPCAVLGHS 120

Qy        121 LGEIAAAHVSGGLSLADAARVVTLWSQAQTTLAGTGALVSVAATPDELLPRIAPWTEDNP 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 LGEIAAAHVSGGLSLADAARVVTLWSQAQTTLAGTGALVSVAATPDELLPRIAPWTEDNP 180

Qy        181 ARLAVAAVNGPRSTVVSGAREAVADLVADLTAAQVRTRMIPVDVPAHSPLMYAIEERVVS 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 ARLAVAAVNGPRSTVVSGAREAVADLVADLTAAQVRTRMIPVDVPAHSPLMYAIEERVVS 240

Qy        241 GLLPITPRPSRIPFHSSVTGGRLDTRELDAAYWYRNMSSTVRFEPAARLLLQQGPKTFVE 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GLLPITPRPSRIPFHSSVTGGRLDTRELDAAYWYRNMSSTVRFEPAARLLLQQGPKTFVE 300

Qy        301 MSPHPVLTMGLQELAPDLGDTTGTADTVIMGTLRRGQGTLDHFLTSLAQLRGHGETSATT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 MSPHPVLTMGLQELAPDLGDTTGTADTVIMGTLRRGQGTLDHFLTSLAQLRGHGETSATT 360

Qy        361 VLSARLTALSPTQQQSLLLDLVRAHTMAVLNDDGNERTASDAGPSASFAHLGFDSVMGVE 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 VLSARLTALSPTQQQSLLLDLVRAHTMAVLNDDGNERTASDAGPSASFAHLGFDSVMGVE 420

Qy        421 LRNRLSKATGLRLPVTLIFDHTTPAAVAARLRTAALGHLDEDTAPVPDSPSGHGGTAAAD 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 LRNRLSKATGLRLPVTLIFDHTTPAAVAARLRTAALGHLDEDTAPVPDSPSGHGGTAAAD 480

Qy        481 DPIAIIGMACRFPGGVRSPKDLWELAASGGDAIGPFPTDRGWPTEQRHAQDPTQPGTFYP 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 DPIAIIGMACRFPGGVRSPKDLWELAASGGDAIGPFPTDRGWPTEQRHAQDPTQPGTFYP 540

Qy        541 QGGGFLHDAAHFDAGFFGISPREALAMDPQQRLLLETSWEAFERAGIDPLSVRGSRTGVF 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 QGGGFLHDAAHFDAGFFGISPREALAMDPQQRLLLETSWEAFERAGIDPLSVRGSRTGVF 600

Qy        601 AGALSFDYGPRMDTASSEGAADVEGHILTGTTGSVLSGRIAYSFGLEGPAITVDTGCSAS 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 AGALSFDYGPRMDTASSEGAADVEGHILTGTTGSVLSGRIAYSFGLEGPAITVDTGCSAS 660

Qy        661 LVTLHLACQSLRSGECTLALAGGVSVMSTLGMFIEFSRQRGLSVDGRCKAYSAAADGTGW 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 LVTLHLACQSLRSGECTLALAGGVSVMSTLGMFIEFSRQRGLSVDGRCKAYSAAADGTGW 720

Qy        721 GEGVGMLLVERLSDAVRLGHRVLAVVRGSAVNQDGASNGLTAPNGPAQERVIRQALANAG 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GEGVGMLLVERLSDAVRLGHRVLAVVRGSAVNQDGASNGLTAPNGPAQERVIRQALANAG 780

Qy        781 LSVADVDVVEGHGTGTTLGDPIEAQALLATYGQRAGDRPLWLGSLKSNIGHTMAAAGVGG 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 LSVADVDVVEGHGTGTTLGDPIEAQALLATYGQRAGDRPLWLGSLKSNIGHTMAAAGVGG 840

Qy        841 VIKMVMALREGVLPRTLHVDKPSPQVDWSAGAVRLLTEAVPWPGDAAGRLRRAGVSSFGI 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 VIKMVMALREGVLPRTLHVDKPSPQVDWSAGAVRLLTEAVPWPGDAAGRLRRAGVSSFGI 900

Qy        901 GGTNAHVILEEAPAAGGCVAGGGVLEGAPGLAISVAESVAAPVAVSAPVAESVPVPVPVP 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 GGTNAHVILEEAPAAGGCVAGGGVLEGAPGLAISVAESVAAPVAVSAPVAESVPVPVPVP 960

Qy        961 VPVPVSARSEAGLRAQAEALRQYVAVRPDVSLADVGAGLACGRAVLEHRAVVLAADREEL 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 VPVPVSARSEAGLRAQAEALRQYVAVRPDVSLADVGAGLACGRAVLEHRAVVLAADREEL 1020

Qy       1021 VQGLGALAAGEPDRRVTTGHAPGGDRGGVVFVFPGQGGQWAGMGVRLLASSPVFARRMQA 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 VQGLGALAAGEPDRRVTTGHAPGGDRGGVVFVFPGQGGQWAGMGVRLLASSPVFARRMQA 1080

Qy       1081 CEEALAPWVDWSVVDILRRDAGDAVWERADVVQPVLFSVMVSLAALWRSYGIEPDAVLGH 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 CEEALAPWVDWSVVDILRRDAGDAVWERADVVQPVLFSVMVSLAALWRSYGIEPDAVLGH 1140

Qy       1141 SQGEIAAAHVCGALSLKDAAKTVALRSRALAAVRGRGGMASVPLPAQEVEQLIGERWAGR 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 SQGEIAAAHVCGALSLKDAAKTVALRSRALAAVRGRGGMASVPLPAQEVEQLIGERWAGR 1200

Qy       1201 LWVAAVNGPRSTAVSGDAEAVDEVLAYCAGTGVRARRIPVDYASHCPHVQPLREELLELL 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 LWVAAVNGPRSTAVSGDAEAVDEVLAYCAGTGVRARRIPVDYASHCPHVQPLREELLELL 1260

Qy       1261 GDISPQPSGVPFFSTVEGTWLDTTTLDAAYWYRNLHQPVRFSDAVQALADDGHRVFVEVS 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 GDISPQPSGVPFFSTVEGTWLDTTTLDAAYWYRNLHQPVRFSDAVQALADDGHRVFVEVS 1320

Qy       1321 PHPTLVPAIEDTTEDTAEDVTAIGSLRRGDNDTRRFLTALAHTHTTGIGTPTTWHHHYTH 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 PHPTLVPAIEDTTEDTAEDVTAIGSLRRGDNDTRRFLTALAHTHTTGIGTPTTWHHHYTH 1380

Qy       1381 HHTHPHPHTHLDLPTYPFQHQHYWLESSQPGAGSGSGAGAGSGAGSGRAGTAGGTAEVES 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 HHTHPHPHTHLDLPTYPFQHQHYWLESSQPGAGSGSGAGAGSGAGSGRAGTAGGTAEVES 1440
```

```
Qy      1441 RFWDAVARQDLETVATTLAVPPSAGLDTVVPALSAWHRHQHDQARINTWTYQETWKPLTL 1500
Db      1441 RFWDAVARQDLETVATTLAVPPSAGLDTVVPALSAWHRHQHDQARINTWTYQETWKPLTL 1500

Qy      1501 PTTHQPHQTWLIAIPETQTHHPHITNILTNLHHHGITPIPLTLNHTHTNPQHLHHTLHHT 1560
Db      1501 PTTHQPHQTWLIAIPETQTHHPHITNILTNLHHHGITPIPLTLNHTHTNPQHLHHTLHHT 1560

Qy      1561 RQQAQNHTTGAITGLLSLLALDETPHPHHPHTPTGTLLNLTLTQTHTQTHPPTPLWYATT 1620
Db      1561 RQQAQNHTTGAITGLLSLLALDETPHPHHPHTPTGTLLNLTLTQTHTQTHPPTPLWYATT 1620

Qy      1621 NATTTHPNDPLTHPTQAQTWGLARTTLLEHPTHTAGIIDLPTTPTPHTLQHLTQTLTQPH 1680
Db      1621 NATTTHPNDPLTHPTQAQTWGLARTTLLEHPTHTAGIIDLPTTPTPHTLQHLTQTLTQPH 1680

Qy      1681 HQTQLAIRTTGTHTRRLTPTTLTPTHQPPTPTPHGTTLITGGTGALATHLTHHLTTHQPT 1740
Db      1681 HQTQLAIRTTGTHTRRLTPTTLTPTHQPPTPTPHGTTLITGGTGALATHLTHHLTTHQPT 1740

Qy      1741 QHLLLTSRTGPHTPHAQHLTTQLQQKGIHLTITTCDTSNPDQLQQLLNTIPPQHPLTTVI 1800
Db      1741 QHLLLTSRTGPHTPHAQHLTTQLQQKGIHLTITTCDTSNPDQLQQLLNTIPPQHPLTTVI 1800

Qy      1801 HTAGILDDATLTNLTPTQLNNVLRAKAHSAHLLHQLTQHTPLTAFVLYSSAAATFGAPGQ 1860
Db      1801 HTAGILDDATLTNLTPTQLNNVLRAKAHSAHLLHQLTQHTPLTAFVLYSSAAATFGAPGQ 1860

Qy      1861 ANYAAANAYLDALAHHRHTHHLPATSIAWGTWQGNGLADSDKARAYLDRRGFRPMSPELA 1920
Db      1861 ANYAAANAYLDALAHHRHTHHLPATSIAWGTWQGNGLADSDKARAYLDRRGFRPMSPELA 1920

Qy      1921 TAAVTQAIADTERPYVVIADIDWSKIEHTSQTSDLVSAARERERPAVQRPTPPAELHKTLA 1980
Db      1921 TAAVTQAIADTERPYVVIADIDWSKIEHTSQTSDLVSAARERERPAVQRPTPPAELHKTLA 1980

Qy      1981 HQTSADQRAALLELVRDHVAAVLRHADPKAIAPDQSFRALGFDSLTAVEFRNLLIKATGL 2040
Db      1981 HQTSADQRAALLELVRDHVAAVLRHADPKAIAPDQSFRALGFDSLTAVEFRNLLIKATGL 2040

Qy      2041 RLPVSLVFDHPTPAKLAVHLQNQLRGTAAESAPSAAAVTAEASVTEPIAIVGMACRFPGG 2100
Db      2041 RLPVSLVFDHPTPAKLAVHLQNQLRGTAAESAPSAAAVTAEASVTEPIAIVGMACRFPGG 2100

Qy      2101 VTSADDFWDLISSEQDAIGGFPTDRGWDLDTLYDPDPDHPGTCYTRNGGFLYDAGHFDAE 2160
Db      2101 VTSADDFWDLISSEQDAIGGFPTDRGWDLDTLYDPDPDHPGTCYTRNGGFLYDAGHFDAE 2160

Qy      2161 FFGISPREALAMDPQQRLLLETAWETIEHAGINPHTLHGTPTGVFTGTNGQDYALRVHNA 2220
Db      2161 FFGISPREALAMDPQQRLLLETAWETIEHAGINPHTLHGTPTGVFTGTNGQDYALRVHNA 2220

Qy      2221 GQSTDGFALTGTAGSVISGRISYTFGFEGPAVSVDTACSSSLVALHLACQALRAGECSMA 2280
Db      2221 GQSTDGFALTGTAGSVISGRISYTFGFEGPAVSVDTACSSSLVALHLACQALRAGECSMA 2280

Qy      2281 LAGGVTVMSSPGAFVEFSRQRGLAADGHCKAFSAAADGTGWGEGVGMLLVERLSDAHRNG 2340
Db      2281 LAGGVTVMSSPGAFVEFSRQRGLAADGHCKAFSAAADGTGWGEGVGMLLVERLSDAHRNG 2340

Qy      2341 HRVLAVVRGSAVNQDGASNGLTAPNGPSQQRVIRQALANAGLSAGDVDAVEAHGTGTTLG 2400
Db      2341 HRVLAVVRGSAVNQDGASNGLTAPNGPSQQRVIRQALANAGLSAGDVDAVEAHGTGTTLG 2400

Qy      2401 DPIEAQALLATYGQDRAGEGPLWLGSVKSNVGHTQAAAGVAGVIKMVMALRHGLLPRTLH 2460
Db      2401 DPIEAQALLATYGQDRAGEGPLWLGSVKSNVGHTQAAAGVAGVIKMVMALRHGLLPRTLH 2460

Qy      2461 VDEPSPHVDWSAGAVQLLTETVPWPGGEGRLRRAGVSSFGVSGTNAHVILEEAPADDVPG 2520
Db      2461 VDEPSPHVDWSAGAVQLLTETVPWPGGEGRLRRAGVSSFGVSGTNAHVILEEAPADDVPG 2520

Qy      2521 GPPAGEGDAGSDDEAAAGSPGVWPWLVSAKSQPALRAQAQALHAHLTDHPGLDLADVGYT 2580
Db      2521 GPPAGEGDAGSDDEAAAGSPGVWPWLVSAKSQPALRAQAQALHAHLTDHPGLDLADVGYT 2580

Qy      2581 LAHARAVFDHRATLIAADRDTFLQALQALAAGEPHPAVIHSSAPGGTGTGEAAGKTAFIC 2640
Db      2581 LAHARAVFDHRATLIAADRDTFLQALQALAAGEPHPAVIHSSAPGGTGTGEAAGKTAFIC 2640

Qy      2641 SGQGTQRPGMAHGLYHTHPVFAAALNDICTHLDPHLDHPLLPLLTQNDNDNEDAAALLQQ 2700
Db      2641 SGQGTQRPGMAHGLYHTHPVFAAALNDICTHLDPHLDHPLLPLLTQNDNDNEDAAALLQQ 2700

Qy      2701 TRYAQPALFAFQVALHRLLTDGYHITPHYYAGHSLGEITAAHLAGILTLTDATTLITQRA 2760
Db      2701 TRYAQPALFAFQVALHRLLTDGYHITPHYYAGHSLGEITAAHLAGILTLTDATTLITQRA 2760

Qy      2761 TLMQTMPPGTMTTLHTTPHHITHHLTAHENDLAIAAINTPTSLVISGTPHTVQHITTLCQ 2820
Db      2761 TLMQTMPPGTMTTLHTTPHHITHHLTAHENDLAIAAINTPTSLVISGTPHTVQHITTLCQ 2820

Qy      2821 QQGIKTKTLPTNHAFHSPHTNPILNQLHQHTQTLTYHPPHTPLITANTPPDQLLTPHYWT 2880
Db      2821 QQGIKTKTLPTNHAFHSPHTNPILNQLHQHTQTLTYHPPHTPLITANTPPDQLLTPHYWT 2880

Qy      2881 QQARNTVDYATTTQTLHQHGVTTYIELGPDNTLTTLTHHNLPNPPTTTLTLTHPHHHPQT 2940
Db      2881 QQARNTVDYATTTQTLHQHGVTTYIELGPDNTLTTLTHHNLPNPPTTTLTLTHPHHHPQT 2940

Qy      2941 HLLTNLAKTTTTWHPHHYTHHDNQPHTHTHLDLPTYPFQHHHYWLESTQPGAGNVSAAGL 3000
Db      2941 HLLTNLAKTTTTWHPHHYTHHDNQPHTHTHLDLPTYPFQHHHYWLESTQPGAGNVSAAGL 3000

Qy      3001 DPTEHPLLGATLELATDGGALLAGRLSLRSHPWLADHAVGGTVLLSGATFLELALHAGTY 3060
Db      3001 DPTEHPLLGATLELATDGGALLAGRLSLRSHPWLADHAVGGTVLLSGATFLELALHAGTY 3060

Qy      3061 VGCDRVDELTLHAPLVVPVDGGVSVQVGVAAADGEGRRLVSVYARGGSACGGGGASGGVW 3120
Db      3061 VGCDRVDELTLHAPLVVPVDGGVSVQVGVAAADGEGRRLVSVYARGGSACGGGGASGGVW 3120

Qy      3121 TCHASGVLVEAAAGGVVVDGLAGVWPPRGAVAVDVDGVRDRLAGAGCVLGPVFSGLRAVW 3180
Db      3121 TCHASGVLVEAAAGGVVVDGLAGVWPPRGAVAVDVDGVRDRLAGAGCVLGPVFSGLRAVW 3180

Qy      3181 RDGGDLLAEVCLPEEAWGDAAGFGLHPALLDGVVQPLSVLLPGGTGFGEGAGFGEGVRVP 3240
Db      3181 RDGGDLLAEVCLPEEAWGDAAGFGLHPALLDGVVQPLSVLLPGGTGFGEGAGFGEGVRVP 3240

Qy      3241 AVWGGVSLHRAGVTGVRVRVSAVGRGGGREAVSVVVGDEAGVPVASVDRLELRPVDMGQL 3300
Db      3241 AVWGGVSLHRAGVTGVRVRVSAVGRGGGREAVSVVVGDEAGVPVASVDRLELRPVDMGQL 3300

Qy      3301 RAVSVSAGRRGSLYAVQWAEVGPVPVCGQAWAWHEDVGESGGGPVPGVVVLRCPDAGAGG 3360
Db      3301 RAVSVSAGRRGSLYAVQWAEVGPVPVCGQAWAWHEDVGESGGGPVPGVVVLRCPDAGAGG 3360

Qy      3361 GGGGGGGGGVGEVVGGVLGVVQGWLGLERFAGSRLVVVTRGAVVAGPEDGPVDVVGASVW 3420
Db      3361 GGGGGGGGGVGEVVGGVLGVVQGWLGLERFAGSRLVVVTRGAVVAGPEDGPVDVVGASVW 3420

Qy      3421 GLVRSAQAEHPDRFVLLDLDTDTGTDLDTGAGAGWGVDGGRVAAVVACGEPQLAVRGERL 3480
Db      3421 GLVRSAQAEHPDRFVLLDLDTDTGTDLDTGAGAGWGVDGGRVAAVVACGEPQLAVRGERL 3480

Qy      3481 LAARLKRLESSGDVPAQRSGDTRARRSDVPAQRSGGVPARRSVDVSGREVLPWLSGGSVL 3540
Db      3481 LAARLKRLESSGDVPAQRSGDTRARRSDVPAQRSGGVPARRSVDVSGREVLPWLSGGSVL 3540

Qy      3541 VTGGTGVLGAAVARHLAGVCGVRDLLLVSRRGPDAPGAEGLRAELAALGAEVRIVACDVG 3600
Db      3541 VTGGTGVLGAAVARHLAGVCGVRDLLLVSRRGPDAPGAEGLRAELAALGAEVRIVACDVG 3600

Qy      3601 ERREVVRLLEGVPAGCPLTGVVHAAGVLDDATIASLTPERLGTVFAAKVDAALLLDELTR 3660
```

```
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3601 ERREVVRLLEGVPAGCPLTGVVHAAGVLDDATIASLTPERLGTVFAAKVDAALLLDELTR 3660

Qy       3661 GMELSAFVLFSSAAGILGSAGQGNYAAANAALDALAYRRRAAGLPGVSLAWGLWEEASGM 3720
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3661 GMELSAFVLFSSAAGILGSAGQGNYAAANAALDALAYRRRAAGLPGVSLAWGLWEEASGM 3720

Qy       3721 TGHLAGTDHRRIIRSGLHPMSTPDALALFDAALALDRPVLLPADLRPAPPLPPLLQDLLP 3780
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3721 TGHLAGTDHRRIIRSGLHPMSTPDALALFDAALALDRPVLLPADLRPAPPLPPLLQDLLP 3780

Qy       3781 ATRRRTTRTTTTGGADNGAQLHARLAGQTHEQQHTTLLALVRSHIATVLGHTTPDTIPPD 3840
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3781 ATRRRTTRTTTTGGADNGAQLHARLAGQTHEQQHTTLLALVRSHIATVLGHTTPDTIPPD 3840

Qy       3841 RAFRDLGFDSLTAVELRNRLSRTTGLRLPTTLAFDHPNPTTLTHHLHTQLQPQPDNAVAP 3900
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3841 RAFRDLGFDSLTAVELRNRLSRTTGLRLPTTLAFDHPNPTTLTHHLHTQLQPQPDNAVAP 3900

Qy       3901 VLAELDKLESALSALDKTDSASERVTLRLKSLMLRWNAPQHPTAESADDDEKFTSATEAE 3960
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3901 VLAELDKLESALSALDKTDSASERVTLRLKSLMLRWNAPQHPTAESADDDEKFTSATEAE 3960

Qy       3961 IFKFIDNDLGLS 3972
                ||||||||||||
Db       3961 IFKFIDNDLGLS 3972
```

RESULT 3
AAG65268
ID AAG65268 standard; Protein; 3972 AA.
XX
AC AAG65268;
XX
DT 04-DEC-2001 (first entry)
XX
DE Streptomyces avermitilis protein derivative SEQ ID NO: 8.
XX
KW Avermectin aglycone synthase; AAS; avermectin derivative;
KW drug production; vetinary drug; pesticide.
XX
OS Synthetic.
XX
PN WO200162939-A1.
XX
PD 30-AUG-2001.
XX
PF 23-FEB-2001; 2001WO-JP01381.
XX
PR 24-FEB-2000; 2000JP-0047405.
XX
PA (KYOW) KYOWA HAKKO KOGYO KK.
PA (KITA) KITASATO INST.
XX
PI Endo H, Yamaguchi H, Kanda Y, Hashimoto S, Omura S, Ikeda H;
XX
DR WPI; 2001-582053/65.
DR N-PSDB; AAH79279.
XX
PT New modified avermectin aglycone synthase derived from Streptomyces
PT avermectilis used in production of 22,23-dihydroavermectin B1a used in
PT drugs and pesticides -
XX
PS Claim 5; Page 235-248; 257pp; Japanese.
XX
CC The present invention relates to the production of modified derivatives
CC of avermectin aglycone synthase (AAS) derived from Streptomyces
CC avermitilis. The activity of an acyl carrier protein (ACP),
CC beta-ketoacyl carrier protein synthase (KS), acyltransferase (AT),
CC beta-ketoacyl carrier protein reductase (KR), dehydratase (DH), enoyl
CC reductase (ER) and/or thioesterase (TE) domain may be reduced or
CC suppressed. The process can be used in the production of drugs, vetinary
CC drugs and pesticides. The present sequence is derived from an S.
CC avermitilis protein.
XX
SQ Sequence 3972 AA;

```
  Query Match             99.9%;  Score 20817;  DB 22;  Length 3972;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 3971;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 VQRMDGGEEPRPAAGEVLGVADEADGGVVFVFPGQGPQWPGMGRELLDASDVFRESVRAC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 VQRMDGGEEPRPAAGEVLGVADEADGGVVFVFPGQGPQWPGMGRELLDASDVFRESVRAC 60

Qy         61 EAAFAPYVDWSVEQVLRDSPDAPGLDRVDVVQPTLFAVMISLAALWRSQGVEPCAVLGHS 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 EAAFAPYVDWSVEQVLRDSPDAPGLDRVDVVQPTLFAVMISLAALWRSQGVEPCAVLGHS 120

Qy        121 LGEIAAAHVSGGLSLADAARVVTLWSQAQTTLAGTGALVSVAATPDELLPRIAPWTEDNP 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 LGEIAAAHVSGGLSLADAARVVTLWSQAQTTLAGTGALVSVAATPDELLPRIAPWTEDNP 180

Qy        181 ARLAVAAVNGPRSTVVSGAREAVADLVADLTAAQVRTRMIPVDVPAHSPLMYAIEERVVS 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 ARLAVAAVNGPRSTVVSGAREAVADLVADLTAAQVRTRMIPVDVPAHSPLMYAIEERVVS 240

Qy        241 GLLPITPRPSRIPFHSSVTGGRLDTRELDAAYWYRNMSSTVRFEPAARLLLQQGPKTFVE 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GLLPITPRPSRIPFHSSVTGGRLDTRELDAAYWYRNMSSTVRFEPAARLLLQQGPKTFVE 300

Qy        301 MSPHPVLTMGLQELAPDLGDTTGTADTVIMGTLRRGQGTLDHFLTSLAQLRGHGETSATT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 MSPHPVLTMGLQELAPDLGDTTGTADTVIMGTLRRGQGTLDHFLTSLAQLRGHGETSATT 360

Qy        361 VLSARLTALSPTQQQSLLLDLVRAHTMAVLNDDGNERTASDAGPSASFAHLGFDSVMGVE 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 VLSARLTALSPTQQQSLLLDLVRAHTMAVLNDDGNERTASDAGPSASFAHLGFDSVMGVE 420

Qy        421 LRNRLSKATGLRLPVTLIFDHTTPAAVAARLRTAALGHLDEDTAPVPDSPSGHGGTAAAD 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 LRNRLSKATGLRLPVTLIFDHTTPAAVAARLRTAALGHLDEDTAPVPDSPSGHGGTAAAD 480

Qy        481 DPIAIIGMACRFPGGVRSPKDLWELAASGGDAIGPFPTDRGWPTEQRHAQDPTQPGTFYP 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 DPIAIIGMACRFPGGVRSPKDLWELAASGGDAIGPFPTDRGWPTEQRHAQDPTQPGTFYP 540

Qy        541 QGGGFLHDAAHFDAGFFGISPREALAMDPQQRLLLETSWEAFERAGIDPLSVRGSRTGVF 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 QGGGFLHDAAHFDAGFFGISPREALAMDPQQRLLLETSWEAFERAGIDPLSVRGSRTGVF 600

Qy        601 AGALSFDYGPRMDTASSEGAADVEGHILTGTTGSVLSGRIAYSFGLEGPAITVDTGCSAS 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db        601 AGALSFDYGPRMDTASSEGAADVEGHILTGTTGSVLSGRIAYSFGLEGPAITVDTGGSAS 660

Qy        661 LVTLHLACQSLRSGECTLALAGGVSVMSTLGMFIEFSRQRGLSVDGRCKAYSAAADGTGW 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 LVTLHLACQSLRSGECTLALAGGVSVMSTLGMFIEFSRQRGLSVDGRCKAYSAAADGTGW 720

Qy        721 GEGVGMLLVERLSDAVRLGHRVLAVVRGSAVNQDGASNGLTAPNGPAQERVIRQALANAG 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GEGVGMLLVERLSDAVRLGHRVLAVVRGSAVNQDGASNGLTAPNGPAQERVIRQALANAG 780

Qy        781 LSVADVDVVEGHGTGTTLGDPIEAQALLATYGQRAGDRPLWLGSLKSNIGHTMAAAGVGG 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 LSVADVDVVEGHGTGTTLGDPIEAQALLATYGQRAGDRPLWLGSLKSNIGHTMAAAGVGG 840

Qy        841 VIKMVMALREGVLPRTLHVDKPSPQVDWSAGAVRLLTEAVPWPGDAAGRLRRAGVSSFGI 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 VIKMVMALREGVLPRTLHVDKPSPQVDWSAGAVRLLTEAVPWPGDAAGRLRRAGVSSFGI 900

Qy        901 GGTNAHVILEEAPAAGGCVAGGGVLEGAPGLAISVAESVAAPVAVSAPVAESVPVPVPVP 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 GGTNAHVILEEAPAAGGCVAGGGVLEGAPGLAISVAESVAAPVAVSAPVAESVPVPVPVP 960

Qy        961 VPVPVSARSEAGLRAQAEALRQYVAVRPDVSLADVGAGLACGRAVLEHRAVVLAADREEL 1020
```

```
Db        961 VPVPVSARSEAGLRAQAEALRQYVAVRPDVSLADVGAGLACGRAVLEHRAVVLAADREEL 1020

Qy       1021 VQGLGALAAGEPDRRVTTGHAPGGDRGGVVFVFPGQGGQWAGMGVRLLASSPVFARRMQA 1080

Db       1021 VQGLGALAAGEPDRRVTTGHAPGGDRGGVVFVFPGQGGQWAGMGVRLLASSPVFARRMQA 1080

Qy       1081 CEEALAPWVDWSVVDILRRDAGDAVWERADVVQPVLFSVMVSLAALWRSYGIEPDAVLGH 1140

Db       1081 CEEALAPWVDWSVVDILRRDAGDAVWERADVVQPVLFSVMVSLAALWRSYGIEPDAVLGH 1140

Qy       1141 SQGEIAAAHVCGALSLKDAAKTVALRSRALAAVRGRGGMASVPLPAQEVEQLIGERWAGR 1200

Db       1141 SQGEIAAAHVCGALSLKDAAKTVALRSRALAAVRGRGGMASVPLPAQEVEQLIGERWAGR 1200

Qy       1201 LWVAAVNGPRSTAVSGDAEAVDEVLAYCAGTGVRARRIPVDYASHCPHVQPLREELLELL 1260

Db       1201 LWVAAVNGPRSTAVSGDAEAVDEVLAYCAGTGVRARRIPVDYASHCPHVQPLREELLELL 1260

Qy       1261 GDISPQPSGVPFFSTVEGTWLDTTTLDAAYWYRNLHQPVRFSDAVQALADDGHRVFVEVS 1320

Db       1261 GDISPQPSGVPFFSTVEGTWLDTTTLDAAYWYRNLHQPVRFSDAVQALADDGHRVFVEVS 1320

Qy       1321 PHPTLVPAIEDTTEDTAEDVTAIGSLRRGDNDTRRFLTALAHTHTTGIGTPTTWHHHYTH 1380

Db       1321 PHPTLVPAIEDTTEDTAEDVTAIGSLRRGDNDTRRFLTALAHTHTTGIGTPTTWHHHYTH 1380

Qy       1381 HHTHPHPHTHLDLPTYPFQHQHYWLESSQPGAGSGSGAGAGSGAGSGRAGTAGGTAEVES 1440

Db       1381 HHTHPHPHTHLDLPTYPFQHQHYWLESSQPGAGSGSGAGAGSGAGSGRAGTAGGTAEVES 1440

Qy       1441 RFWDAVARQDLETVATTLAVPPSAGLDTVVPALSAWHRHQHDQARINTWTYQETWKPLTL 1500

Db       1441 RFWDAVARQDLETVATTLAVPPSAGLDTVVPALSAWHRHQHDQARINTWTYQETWKPLTL 1500

Qy       1501 PTTHQPHQTWLIAIPETQTHHPHITNILTNLHHHGITPIPLTLNHTHTNPQHLHHTLHHT 1560

Db       1501 PTTHQPHQTWLIAIPETQTHHPHITNILTNLHHHGITPIPLTLNHTHTNPQHLHHTLHHT 1560

Qy       1561 RQQAQNHTTGAITGLLSLLALDETPHPHHPHTPTGTLLNLTLTQTHTQTHPPTPLWYATT 1620

Db       1561 RQQAQNHTTGAITGLLSLLALDETPHPHHPHTPTGTLLNLTLTQTHTQTHPPTPLWYATT 1620

Qy       1621 NATTTHPNDPLTHPTQAQTWGLARTTLLEHPTHTAGIIDLPTTPTPHTLQHLTQTLTQPH 1680

Db       1621 NATTTHPNDPLTHPTQAQTWGLARTTLLEHPTHTAGIIDLPTTPTPHTLQHLTQTLTQPH 1680

Qy       1681 HQTQLAIRTTGTHTRRLTPTTLTPTHQPPTPTPHGTTLITGGTGALATHLTHHLTTHQPT 1740

Db       1681 HQTQLAIRTTGTHTRRLTPTTLTPTHQPPTPTPHGTTLITGGTGALATHLTHHLTTHQPT 1740

Qy       1741 QHLLLTSRTGPHTPHAQHLTTQLQQKGIHLTITTCDTSNPDQLQQLLNTIPPQHPLTTVI 1800

Db       1741 QHLLLTSRTGPHTPHAQHLTTQLQQKGIHLTITTCDTSNPDQLQQLLNTIPPQHPLTTVI 1800

Qy       1801 HTAGILDDATLTNLTPTQLNNVLRAKAHSAHLLHQLTQHTPLTAFVLYSSAAATFGAPGQ 1860

Db       1801 HTAGILDDATLTNLTPTQLNNVLRAKAHSAHLLHQLTQHTPLTAFVLYSSAAATFGAPGQ 1860

Qy       1861 ANYAAANAYLDALAHHRHTHHLPATSIAWGTWQGNGLADSDKARAYLDRRGFRPMSPELA 1920

Db       1861 ANYAAANAYLDALAHHRHTHHLPATSIAWGTWQGNGLADSDKARAYLDRRGFRPMSPELA 1920

Qy       1921 TAAVTQAIADTERPYVVIADIDWSKIEHTSQTSDLVSAARERERPAVQRPTPPAELHKTLA 1980

Db       1921 TAAVTQAIADTERPYVVIADIDWSKIEHTSQTSDLVSAARERERPAVQRPTPPAELHKTLA 1980

Qy       1981 HQTSADQRAALLELVRDHVAAVLRHADPKAIAPDQSFRALGFDSLTAVEFRNLLIKATGL 2040

Db       1981 HQTSADQRAALLELVRDHVAAVLRHADPKAIAPDQSFRALGFDSLTAVEFRNLLIKATGL 2040

Qy       2041 RLPVSLVFDHPTPAKLAVHLQNQLRGTAAESAPSAAAVTAEASVTEPIAIVGMACRFPGG 2100

Db       2041 RLPVSLVFDHPTPAKLAVHLQNQLRGTAAESAPSAAAVTAEASVTEPIAIVGMACRFPGG 2100

Qy       2101 VTSADDFWDLISSEQDAIGGFPTDRGWDLDTLYDPDPDHPGTCYTRNGGFLYDAGHFDAE 2160

Db       2101 VTSADDFWDLISSEQDAIGGFPTDRGWDLDTLYDPDPDHPGTCYTRNGGFLYDAGHFDAE 2160

Qy       2161 FFGISPREALAMDPQQRLLLETAWETIEHAGINPHTLHGTPTGVFTGTNGQDYALRVHNA 2220

Db       2161 FFGISPREALAMDPQQRLLLETAWETIEHAGINPHTLHGTPTGVFTGTNGQDYALRVHNA 2220

Qy       2221 GQSTDGFALTGTAGSVISGRISYTFGFEGPAVSVDTACSSSLVALHLACQALRAGECSMA 2280

Db       2221 GQSTDGFALTGTAGSVISGRISYTFGFEGPAVSVDTACSSSLVALHLACQALRAGECSMA 2280

Qy       2281 LAGGVTVMSSPGAFVEFSRQRGLAADGHCKAFSAAADGTGWGEGVGMLLVERLSDAHRNG 2340

Db       2281 LAGGVTVMSSPGAFVEFSRQRGLAADGHCKAFSAAADGTGWGEGVGMLLVERLSDAHRNG 2340

Qy       2341 HRVLAVVRGSAVNQDGASNGLTAPNGPSQQRVIRQALANAGLSAGDVDAVEAHGTGTTLG 2400

Db       2341 HRVLAVVRGSAVNQDGASNGLTAPNGPSQQRVIRQALANAGLSAGDVDAVEAHGTGTTLG 2400

Qy       2401 DPIEAQALLATYGQDRAGEGPLWLGSVKSNVGHTQAAAGVAGVIKMVMALRHGLLPRTLH 2460

Db       2401 DPIEAQALLATYGQDRAGEGPLWLGSVKSNVGHTQAAAGVAGVIKMVMALRHGLLPRTLH 2460

Qy       2461 VDEPSPHVDWSAGAVQLLTETVPWPGGEGRLRRAGVSSFGVSGTNAHVILEEAPADDVPG 2520

Db       2461 VDEPSPHVDWSAGAVQLLTETVPWPGGEGRLRRAGVSSFGVSGTNAHVILEEAPADDVPG 2520

Qy       2521 GPPAGEGDAGSDDEAAAGSPGVWPWLVSAKSQPALRAQAQALHAHLTDHPGLDLADVGYT 2580

Db       2521 GPPAGEGDAGSDDEAAAGSPGVWPWLVSAKSQPALRAQAQALHAHLTDHPGLDLADVGYT 2580

Qy       2581 LAHARAVFDHRATLIAADRDTFLQALQALAAGEPHPAVIHSSAPGGTGTGEAAGKTAFIC 2640

Db       2581 LAHARAVFDHRATLIAADRDTFLQALQALAAGEPHPAVIHSSAPGGTGTGEAAGKTAFIC 2640

Qy       2641 SGQGTQRPGMAHGLYHTHPVFAAALNDICTHLDPHLDHPLLPLLTQNDNDNEDAAALLQQ 2700

Db       2641 SGQGTQRPGMAHGLYHTHPVFAAALNDICTHLDPHLDHPLLPLLTQNDNDNEDAAALLQQ 2700

Qy       2701 TRYAQPALFAFQVALHRLLTDGYHITPHYYAGHSLGEITAAHLAGILTLTDATTLITQRA 2760

Db       2701 TRYAQPALFAFQVALHRLLTDGYHITPHYYAGHSLGEITAAHLAGILTLTDATTLITQRA 2760

Qy       2761 TLMQTMPPGTMTTLHTTPHHITHHLTAHENDLAIAAAINTPTSLVISGTPHTVQHITTLCQ 2820

Db       2761 TLMQTMPPGTMTTLHTTPHHITHHLTAHENDLAIAAAINTPTSLVISGTPHTVQHITTLCQ 2820

Qy       2821 QQGIKTKTLPTNHAFHSPHTNPILNQLHQHTQTLTYHPPHTPLITANTPPDQLLTPHYWT 2880

Db       2821 QQGIKTKTLPTNHAFHSPHTNPILNQLHQHTQTLTYHPPHTPLITANTPPDQLLTPHYWT 2880

Qy       2881 QQARNTVDYATTTQTLHQHGVTTYIELGPDNTLTTLTHHNLPNPPTTTLTLTHPHHHPQT 2940

Db       2881 QQARNTVDYATTTQTLHQHGVTTYIELGPDNTLTTLTHHNLPNPPTTTLTLTHPHHHPQT 2940

Qy       2941 HLLTNLAKTTTTWHPHHYTHHDNQPHTHTHLDLPTYPFQHHHYWLESTQPGAGNVSAAGL 3000

Db       2941 HLLTNLAKTTTTWHPHHYTHHDNQPHTHTHLDLPTYPFQHHHYWLESTQPGAGNVSAAGL 3000

Qy       3001 DPTEHPLLGATLELATDGGALLAGRLSLRSHPWLADHAVGGTVLLSGATFLELALHAGTY 3060

Db       3001 DPTEHPLLGATLELATDGGALLAGRLSLRSHPWLADHAVGGTVLLSGATFLELALHAGTY 3060

Qy       3061 VGCDRVDELTLHAPLVVPVDGGVSVQVGVAAADGEGRRLVSVYARGGSACGGGGASGGVW 3120

Db       3061 VGCDRVDELTLHAPLVVPVDGGVSVQVGVAAADGEGRRLVSVYARGGSACGGGGASGGVW 3120

Qy       3121 TCHASGVLVEAAAGGVVVDGLAGVWPPRGAVAVDVDGVRDRLAGAGCVLGPVFSGLRAVW 3180

Db       3121 TCHASGVLVEAAAGGVVVDGLAGVWPPRGAVAVDVDGVRDRLAGAGCVLGPVFSGLRAVW 3180
```

```
Qy       3181 RDGGDLLAEVCLPEEAWGDAAGFGLHPALLDGVVQPLSVLLPGGTGFGEGAGFGEGVRVP 3240
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3181 RDGGDLLAEVCLPEEAWGDAAGFGLHPALLDGVVQPLSVLLPGGTGFGEGAGFGEGVRVP 3240

Qy       3241 AVWGGVSLHRAGVTGVRVRVSAVGRGGGREAVSVVVGDEAGVPVASVDRLELRPVDMGQL 3300
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3241 AVWGGVSLHRAGVTGVRVRVSAVGRGGGREAVSVVVGDEAGVPVASVDRLELRPVDMGQL 3300

Qy       3301 RAVSVSAGRRGSLYAVQWAEVGPVPVCGQAWAWHEDVGESGGGPVPGVVVLRCPDAGAGG 3360
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3301 RAVSVSAGRRGSLYAVQWAEVGPVPVCGQAWAWHEDVGESGGGPVPGVVVLRCPDAGAGG 3360

Qy       3361 GGGGGGGGGVGEVVGGVLGVVQGWLGLERFAGSRLVVVTRGAVVAGPEDGPVDVVGASVW 3420
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3361 GGGGGGGGGVGEVVGGVLGVVQGWLGLERFAGSRLVVVTRGAVVAGPEDGPVDVVGASVW 3420

Qy       3421 GLVRSAQAEHPDRFVLLDLDTDTGTDLDTGAGAGWGVDGGRVAAVVACGEPQLAVRGERL 3480
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3421 GLVRSAQAEHPDRFVLLDLDTDTGTDLDTGAGAGWGVDGGRVAAVVACGEPQLAVRGERL 3480

Qy       3481 LAARLKRLESSGDVPAQRSGDTRARRSDVPAQRSGGVPARRSVDVSGREVLPWLSGGSVL 3540
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3481 LAARLKRLESSGDVPAQRSGDTRARRSDVPAQRSGGVPARRSVDVSGREVLPWLSGGSVL 3540

Qy       3541 VTGGTGVLGAAVARHLAGVCGVRDLLLVSRRGPDAPGAEGLRAELAALGAEVRIVACDVG 3600
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3541 VTGGTGVLGAAVARHLAGVCGVRDLLLVSRRGPDAPGAEGLRAELAALGAEVRIVACDVG 3600

Qy       3601 ERREVVRLLEGVPAGCPLTGVVHAAGVLDDATIASLTPERLGTVFAAKVDAALLLDELTR 3660
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3601 ERREVVRLLEGVPAGCPLTGVVHAAGVLDDATIASLTPERLGTVFAAKVDAALLLDELTR 3660

Qy       3661 GMELSAFVLFSSAAGILGSAGQGNYAAANAALDALAYRRRAAGLPGVSLAWGLWEEASGM 3720
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3661 GMELSAFVLFSSAAGILGSAGQGNYAAANAALDALAYRRRAAGLPGVSLAWGLWEEASGM 3720

Qy       3721 TGHLAGTDHRRIIRSGLHPMSTPDALALFDAALALDRPVLLPADLRPAPPLPPLLQDLLP 3780
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3721 TGHLAGTDHRRIIRSGLHPMSTPDALALFDAALALDRPVLLPADLRPAPPLPPLLQDLLP 3780

Qy       3781 ATRRRTTRTTTTGGADNGAQLHARLAGQTHEQQHTTLLALVRSHIATVLGHTTPDTIPPD 3840
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3781 ATRRRTTRTTTTGGADNGAQLHARLAGQTHEQQHTTLLALVRSHIATVLGHTTPDTIPPD 3840

Qy       3841 RAFRDLGFDSLTAVELRNRLSRTTGLRLPTTLAFDHPNPTTLTHHLHTQLQPQPDNAVAP 3900
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3841 RAFRDLGFDSLTAVELRNRLSRTTGLRLPTTLAFDHPNPTTLTHHLHTQLQPQPDNAVAP 3900

Qy       3901 VLAELDKLESALSALDKTDSASERVTLRLKSLMLRWNAPQHPTAESADDDEKFTSATEAE 3960
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3901 VLAELDKLESALSALDKTDSASERVTLRLKSLMLRWNAPQHPTAESADDDEKFTSATEAE 3960

Qy       3961 IFKFIDNDLGLS 3972
             ||||||||||||
Db       3961 IFKFIDNDLGLS 3972
```

```
RESULT 4
AAB23752
ID   AAB23752 standard; Protein; 5532 AA.
XX
AC   AAB23752;
XX
DT   10-JAN-2001  (first entry)
XX
DE   S. avermitilis avermectin aglycon synthase protein SEQ ID NO:6.
XX
KW   Streptomyces avermitilis; avermectin aglycon synthase; biosynthesis;
KW   multifunctional enzyme; polyketide; avermectin; veterinary drug;
KW   agrochemical.
XX
OS   Streptomyces avermitilis.
XX
PN   WO200050605-A1.
XX
PD   31-AUG-2000.
XX
PF   23-FEB-2000; 2000WO-JP01041.
XX
PR   24-FEB-1999;   99JP-0046961.
XX
PA   (KITA ) KITASATO INST.
XX
PI   Omura S,  Ikeda H;
XX
DR   WPI; 2000-565458/52.
DR   N-PSDB; AAA92302.
XX
PT   Avermectin aglycone synthase DNA and proteins encoded by all or part of
PT   it for the production of avermectin and its derivatives for drug and
PT   agrochemical use  -
XX
PS   Claim 32; Page 278-305; 314pp; Japanese.
XX
CC   The present sequence represents an avermectin aglycon synthase protein.
CC   Also described are: (1) polypeptides encoded by all or part of the DNA;
CC   (2) expression vectors containing the DNA; (3) host cells transformed by
CC   the vectors; (4) preparation of the polypeptides by culture of the
CC   transformants; (5) preparation of avermectin aglycon or its derivatives
CC   by culture of transformed avermectin-producing microorganisms; and (6)
CC   oligonucleotides of 5-60 bases in length containing sense or antisense
CC   sequences from the avermectin aglycon synthase DNA. The enzymes are
CC   useful for the production of modified forms of avermectin and of the
CC   intermediates in its biosynthesis, for use as drugs, veterinary drugs
CC   and agrochemicals.
XX
SQ   Sequence   5532 AA;
```

Query Match 62.3%; Score 12972; DB 21; Length 5532;
Best Local Similarity 71.5%; Pred. No. 0;
Matches 2640; Conservative 168; Mismatches 519; Indels 366; Gaps 47;

```
Qy        478 AADDPIAIIGMACRFPGGVRSPKDLWELAASGGDAIGPFPTDRGWPTEQRHAQDPTQPGT 537
              |: :||||:|||||||||| ||::|||| |||||||| ||  |||  |     || : ||
Db         35 ASSEPIAIVGMACRFPGGVCSPEELWELVASGGDAIGEFPAGRGWDLEGLFDSDPDRSGT 94

Qy        538 FYPQGGGFLHDAAHFDAGFFGISPREALAMDPQQRLLLETSWEAFERAGIDPLSVRGSRT 597
               | : ||||::|  ||| ||||||||||||||||||||||||||||||:|||||
Db         95 SYARYGGFLYEAGEFDADFFGISPREALAMDPQQRLLLETSWEAFERAGIDPLSMRGSRT 154

Qy        598 GVFAGALSFDYGPRMDTASSEGAADVEGHILTGTTGSVLSGRIAYSFGLEGPAITVDTGC 657
              ||||| :  ||| |: |   ||    ||:|  |: |:| |||:||: |||||::||| |
Db        155 GVFAGVMYHDYGSRLGTI-PEG---FEGYIGNGSGGAVASGRVAYTLGLEGPAVSVDTAC 210

Qy        658 SASLVTLHLACQSLRSGECTLALAGGVSVMSTLGMFIEFSRQRGLSVDGRCKAYSAAADG 717
              |:||| |||||||||||||||||:|||  :|:|||||||||||::  |||
Db        211 SSSLVALHLACQSLRSGECTLALAGGVTVMSTPHLFVEFSRQRGLSVDGRCKSFAGGADG 270

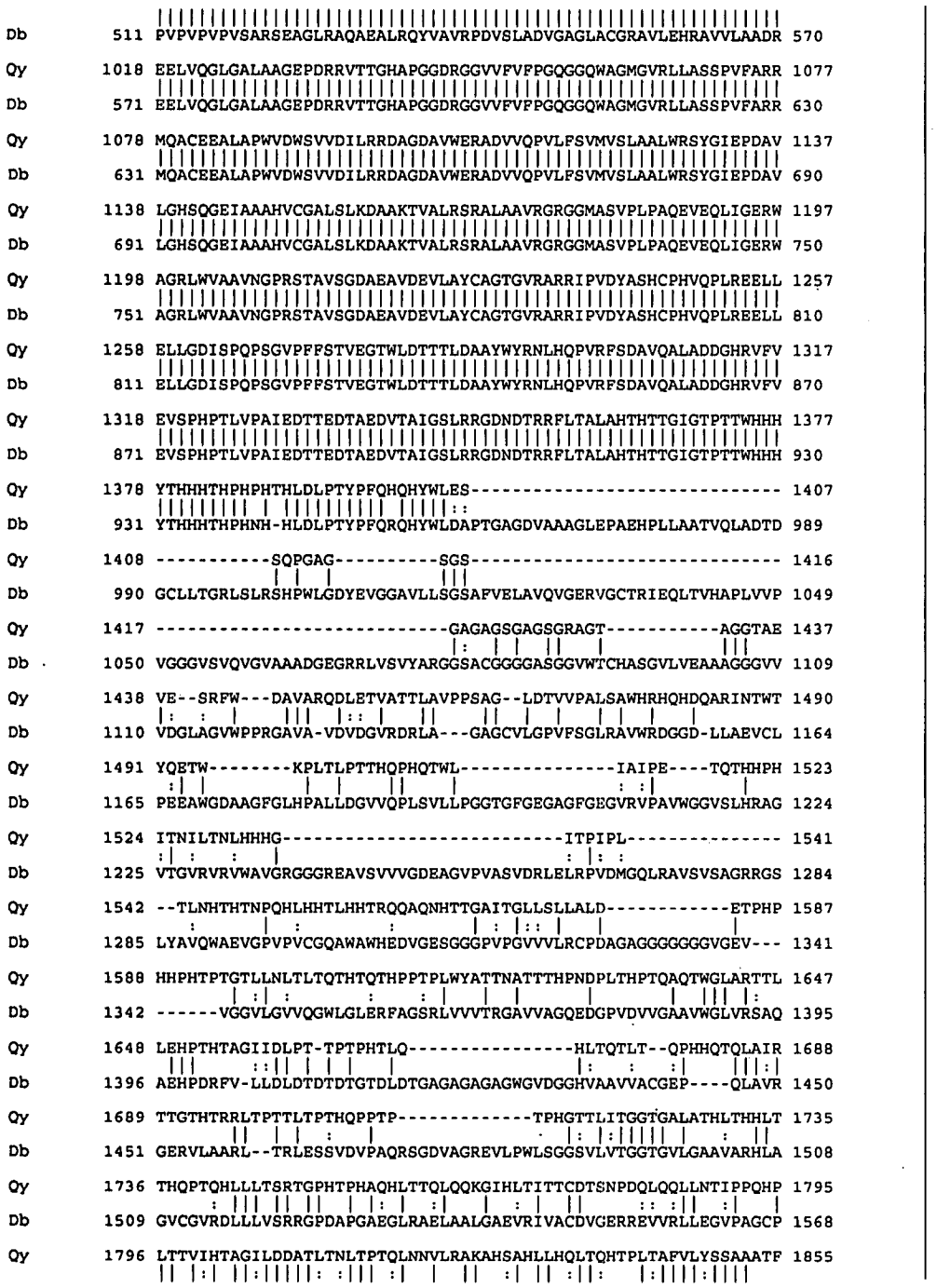
Qy        718 TGWGEGVGMLLVERLSDAVRLGHRVLAVVRGSAVNQDGASNGLTAPNGPAQERVIRQALA 777
              || |||||||||||||||||||:|||||||||||||||||||||||||||||||||||||
Db        271 TGMGEGVGMLLVERLSDAVRLGHRVLAVLRGSAVNQDGASNGLTAPNGPAQERVIRQALA 330

Qy        778 NAGLSVADVDVVEGHGTGTTLGDPIEAQALLATYGQRAGDRPLWLGSLKSNIGHTMAAAG 837
              ||||||||||||||||||||||||||||||||||||:|||||:||||||  ||||
Db        331 NAGLSVADVDVVEGHGTGTTLGDPIEAQALLATYGQRAGNRPLWLGSVKSNIGHAQAAAG 390

Qy        838 VGGVIKMVMALREGVLPRTLHVDKPSPQVDWSAGAVRLLTEAVPWPGDAAGRLRRAGVSS 897
              ||||||||||||||||||||:|||||||||||||||||||||||||||||||||||||||
Db        391 VGGVIKMVMALREGVLPRTLHVDEPSPQVDWSAGAVRLLTEAVPWPGDAAGRLRRAGVSS 450

Qy        898 FGIGGTNAHVILEEAPAAGGCVAGGGVLEGAPGLAISVAESVAAPVAVSAPVAESVPVPV 957
              ||: ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        451 FGVSGTNAHVILEEAPAAGGCVAGGGVLEGAPGLAISVAESVAAPVAVSAPVAESVPVPV 510

Qy        958 PVPVPVPVSARSEAGLRAQAEALRQYVAVRPDVSLADVGAGLACGRAVLEHRAVVLAADR 1017
```


Db 511 PVPVPVPVSARSEAGLRAQAEALRQYVAVRPDVSLADVGAGLACGRAVLEHRAVVLAADR 570
Qy 1018 EELVQGLGALAAGEPDRRVTTGHAPGGDRGGVVFVFPGQGGQWAGMGVRLLASSPVFARR 1077
Db 571 EELVQGLGALAAGEPDRRVTTGHAPGGDRGGVVFVFPGQGGQWAGMGVRLLASSPVFARR 630
Qy 1078 MQACEEALAPWVDWSVVDILRRDAGDAVWERADVVQPVLFSVMVSLAALWRSYGIEPDAV 1137
Db 631 MQACEEALAPWVDWSVVDILRRDAGDAVWERADVVQPVLFSVMVSLAALWRSYGIEPDAV 690
Qy 1138 LGHSQGEIAAAHVCGALSLKDAAKTVALRSRALAAVRGRGGMASVPLPAQEVEQLIGERW 1197
Db 691 LGHSQGEIAAAHVCGALSLKDAAKTVALRSRALAAVRGRGGMASVPLPAQEVEQLIGERW 750
Qy 1198 AGRLWVAAVNGPRSTAVSGDAEAVDEVLAYCAGTGVRARRIPVDYASHCPHVQPLREELL 1257
Db 751 AGRLWVAAVNGPRSTAVSGDAEAVDEVLAYCAGTGVRARRIPVDYASHCPHVQPLREELL 810
Qy 1258 ELLGDISPQPSGVPFFSTVEGTWLDTTTLDAAYWYRNLHQPVRFSDAVQALADDGHRVFV 1317
Db 811 ELLGDISPQPSGVPFFSTVEGTWLDTTTLDAAYWYRNLHQPVRFSDAVQALADDGHRVFV 870
Qy 1318 EVSPHPTLVPAIEDTTEDTAEDVTAIGSLRRGDNDTRRFLTALAHTHTTGIGTPTTWHHH 1377
Db 871 EVSPHPTLVPAIEDTTEDTAEDVTAIGSLRRGDNDTRRFLTALAHTHTTGIGTPTTWHHH 930
Qy 1378 YTHHHTHPHPHTHLDLPTYPFQHQHYWLES---------- 1407
Db 931 YTHHHTHPHNH-HLDLPTYPFQRQHYWLDAPTGAGDVAAAGLEPAEHPLLAATVQLADTD 989
Qy 1408 -----SQPGAG----------SGS---------- 1416
Db 990 GCLLTGRLSLRSHPWLGDYEVGGAVLLSGSAFVELAVQVGERVGCTRIEQLTVHAPLVVP 1049
Qy 1417 ----------GAGAGSGAGSGRAGT----------AGGTAE 1437
Db 1050 VGGGVSVQVGVAAADGEGRRLVSVYARGGSACGGGGASGGVWTCHASGVLVEAAAGGGVV 1109
Qy 1438 VE--SRFW---DAVARQDLETVATTLAVPPSAG--LDTVVPALSAWHRHQHDQARINTWT 1490
Db 1110 VDGLAGVWPPRGAVA-VDVDGVRDRLA---GAGCVLGPVFSGLRAVWRDGGD-LLAEVCL 1164
Qy 1491 YQETW--------KPLTLPTTHQPHQTWL----------IAIPE----TQTHHPH 1523
Db 1165 PEEAWGDAAGFGLHPALLDGVVQPLSVLLPGGTGFGEGAGFGEGVRVPAVWGGVSLHRAG 1224
Qy 1524 ITNILTNLHHHG----------ITPIPL---------- 1541
Db 1225 VTGVRVRVWAVGRGGGREAVSVVVGDEAGVPVASVDRLELRPVDMGQLRAVSVSAGRRGS 1284
Qy 1542 --TLNHTHTNPQHLHHTLHHTRQQAQNHTTGAITGLLSLLALD----------ETPHP 1587
Db 1285 LYAVQWAEVGPVPVCGQAWAWHEDVGESGGGPVPGVVVLRCPDAGAGGGGGGGVGEV--- 1341
Qy 1588 HHPHTPTGTLLNLTLTQTHTQTHPPTPLWYATTNATTTHPNDPLTHPTQAQTWGLARTTL 1647
Db 1342 ------VGGVLGVVQGWLGLERFAGSRLVVVTRGAVVAGQEDGPVDVVGAAVWGLVRSAQ 1395
Qy 1648 LEHPTHTAGIIDLPT-TPTPHTLQ----------HLTQTLT--QPHHQTQLAIR 1688
Db 1396 AEHPDRFV-LLDLDTDTDTGTDLDTGAGAGAGAGWGVDGGHVAAVVACGEP----QLAVR 1450
Qy 1689 TTGTHTRRLTPTTLTPTHQPPTP----------TPHGTTLITGGTGALATHLTHHLT 1735
Db 1451 GERVLAARL--TRLESSVDVPAQRSGDVAGREVLPWLSGGSVLVTGGTGVLGAAVARHLA 1508
Qy 1736 THQPTQHLLLTSRTGPHTPHAQHLTTQLQQKGIHLTITTCDTSNPDQLQQLLNTIPPQHP 1795
Db 1509 GVCGVRDLLLVSRRGPDAPGAEGLRAELAALGAEVRIVACDVGERREVVRLLEGVPAGCP 1568
Qy 1796 LTTVIHTAGILDDATLTNLTPTQLNNVLRAKAHSAHLLHQLTQHTPLTAFVLYSSAAATF 1855
Db 1569 LTGVVHAAGVLDDATIASLTPERLGTVFAAKVDAALLLDELTRGMELSAFVLFSSAAGIL 1628
Qy 1856 GAPGQANYAAANAYLDALAHHRHTHHLPATSIAWGTW-QGNG----LADSDKARAYLDRR 1910
Db 1629 GSAGQGNYAAANAALDALAYRRRAAGLPGVSLAWGLWEEASGMTGHLAGTDHRR--IIRS 1686
Qy 1911 GFRPMSPELATAAVTQAIADTERPYVVIADIDWSKIEHTSQTSDLVSAARAREPAVQRPT 1970
Db 1687 GLHPMSTPDALALFDAALA-LDRPVLLPADL--------RPAPPLPPLLQDLLPATRRRT 1737
Qy 1971 P----------PAELHKTLAHQTSADQRAALLELVRDHVAAVLRHADPKAIAPDQSFRAL 2020
Db 1738 TRTTTTGGADNGAQLHARLAGQTHEQQHTTLLALVRSHIATVLGHTTPDTIPPDRAFRDL 1797
Qy 2021 GFDSLTAVEFRNLLIKATGLRLPVSLVFDHPTPAKLAVHLQNQLRG---TAAESAPSAAA 2077
Db 1798 GFDSLTAVELRNRLSRTTGLRLPTTLAFDHPNPTTLTHHLHTQLLGSDSTASIPAPRAAA 1857
Qy 2078 VTAEASVTEPIAIVGMACRFPGGVTSADDFWDLISSEQDAIGGFPTDRGWDLDTLYDPDP 2137
Db 1858 VPADQD--EPVAIIGMACRYPGGVTSAEELWELLASGRDTVGEFPTDRGWDLEALFDPEP 1915
Qy 2138 DHPGTCYTRNGGFLYDAGHFDAEFFGISPREALAMDPQQRLLLETAWETIEHAGINPHTL 2197
Db 1916 GRPGTSYTRCGSFLYDAGEFDAGFFGISPREALAMDPQQRLLLEASWEAMEQAGIDPTTV 1975
Qy 2198 HGTPTGVFTGTNGQDYALRVH-NAGQSTDGFALTGTAGSVISGRISYTFGFEGPAVSVDT 2256
Db 1976 RGSQTGVFAGLIPQAYGPRLHENAAADTEGYVLTGTSGSVASGRISYTFGFEGPAVSVDT 2035
Qy 2257 ACSSSLVALHLACQALRAGECSMALAGGVTVMSSPGAFVEFSRQRGLAADGHCKAFSAAA 2316
Db 2036 ACSSSLVALHLACQALRAGECSMALAGGVTVMSSPGAFVEFSRQRGLAADGHCKAFSAAA 2095
Qy 2317 DGTGWGEGVGMLLVERLSDAHRNGHRVLAVVRGSAVNQDGASNGLTAPNGPSQQRVIRQA 2376
Db 2096 DGTGWGEGVGMLLVERLSDARRNGHRVLAVVRGSAVNQDGASNGLTAPNGPSQQRVIRQA 2155
Qy 2377 LANAGLSAGDVDAVEAHGTGTTLGDPIEAQALLATYGQDRAGEGPLWLGSVKSNVGHTQA 2436
Db 2156 LANAGLSAGDVDAVEAHGTGTTLGDPIEAQALLATYGQDRAGEGPLWLGSVKSNVGHTQA 2215
Qy 2437 AAGVAGVIKMVMALRHGLLPRTLHVDEPSPHVDWSAGAVQLLTETVPWPGGEGRLRRAGV 2496
Db 2216 AAGVAGVIKMVMALRNGLLPRTLHVDEPSPHVDWSAGAVQLLTETVPWPGGEGRLRRAGV 2275
Qy 2497 SSFGVSGTNAHVILEEA----------PADDVPGGPPAGEGDAGSDDEAAAGSPGVWPWLV 2547
Db 2276 SSFGVSGTNAHVILEEAPAHNIPSDTPADDVPGGPPAGE-DAGSGEEAAAGSPGVWPWLV 2334
Qy 2548 SAKSQPALRAQAQALHAHLTDHPGLDLADVGYTLAHARAVFDHRATLIAADRDTFLQALQ 2607
Db 2335 SAKSQPALRAQAQALHAHLTDHPGLDLADVGYTLAHARAVFDHRATLIAADRDTFLQALQ 2394
Qy 2608 ALAAGEPHPAVIHSSAPGGTGTGEAAGKTAFICSGQGTQRPGMAHGLYHTHPVFAAALND 2667
Db 2395 ALAAGEPHPAVIHSSAPGGTGTGEAAGKTAFICSGQGTQRPGMAHGLYHTHPVFAAALND 2454
Qy 2668 ICTHLDPHLDHPLLPLLTQNDNDN-----EDAAALLQQTRYAQPALFAFQVALHRLLTDG 2722
Db 2455 ICTHLDPHLDHPLLPLLTQDPNTQDTTTLEEAAALLQQTPYAQPALFAFQVALHRLLTDG 2514
Qy 2723 YHITPHYYAGHSLGEITAAHLAGILTLTDATTLITQRATLMQTMPPGTMTTLHTTPHHIT 2782
Db 2515 YHITPHYYAGHSLGEITAAHLAGILTLTDATTLITQRATLMQTMPPGTMTTLHTTPHHIT 2574
Qy 2783 HHLTAHENDLAIAAINTPTSLVISGTPHTVQHITTLCQQQGIKTKTLPTNHAFHSPHTNP 2842
Db 2575 HHITAHENDLAIAAINTPTSLVISGTPHTVQHITTLCQQQGIKTKTLPTNHAFHSPHTNP 2634
Qy 2843 ILNQLHQHTQTLTYHPPHTPLITANTPPDQLLTPHYWTQQARNTVDYATTTQTLHQHGVT 2902
Db 2635 ILNQLHQHTQTLTYHPPHTPLITANTPPDQLLTPHYWTQQARNTVDIATTTQTLHQHGVT 2694

```
Qy       2903 TYIELGPDNTLTTLTHHNLPNPPTTTLTLTHPHHHPQTHLLTNLAKTTTTWHPHHYTHHD 2962
Db       2695 TYIELGPDNTLTTLTHHNLPNTPTTTLTLTHPHHHPQTHLLTNLAKTTTTWHPHHYTHHH 2754

Qy       2963 NQPHTHTHLDLPTYPFQHHHYWLESTQPGAGNVSAAGLDPTEHPLLGATLELATDGGALL 3022
Db       2755 NQPHTHTHLDLPTYPFQHHHYWLESTQPGAGNVSAAGLDPTEHPLLGATLELAEGDGCLL 2814

Qy       3023 AGRLSLRSHPWLADHAVGGTVLLSGATFLELALHAGTYVGCDRVDELTLHAPLVVPVDGG 3082
Db       2815 TGRLSLRTHPWLAGHAVGGVVLLPGTAFAELALHAGESVGCDHVDELTLHTPLVIPEVGD 2874

Qy       3083 VSVQVGVAAADGEGRRLVSVYARGGSACGGGGASGGVWTCHASGVLVEAAAGGVV-VDGL 3141
Db       2875 VTLQVAIAAPDESGRRMMTIHSRGEGGSGGADASASAWTRHAAGVLSPAKDDDTASYELL 2934

Qy       3142 AGVWPPRGAVAVDVDGVRDRLAGAGCVLGPVFSGLRAVWRDGGDLLAEVCLPEEAWGDAA 3201
Db       2935 AGPWPPVGATPVDLNTAYDQMADAGFAYGLAFQGLRAAWRYGDDILVEARLPEEVSGDAA 2994

Qy       3202 GFGLHPALLDGVVQPLSVLLPGGTGFGEGAGFGEGVRVPAVWGGVSLHRAGVTGVRVRVS 3261
Db       2995 AYGLHPALLDAALQ--------GTGLLSVAGPGTPV-VPHVWNGLRFRTHGAVSVRACLS 3045

Qy       3262 AVGRGGGREAVSVVVGDEAGVPVASVDRLELRPVDMGQLRAVSVSAGRRGSLYAVQWAEV 3321
Db       3046 TLGATGA--AVCVRITDDTGVPVASVDRLELRPVDMGQLRAVSVSAGRRGSLYAVQWAEV 3103

Qy       3322 GPVPVCGQAWAWHEDVGESGGGPVPGVVVLRCPDAGAGGGGGGGGGGGVGEVVGGVLGVV 3381
Db       3104 GPVPVCGQAWAWHEDVGESGGGPVPGVVVLRCPDAGA----DGGGGGGVGEVVGGVLGVV 3159

Qy       3382 QGWLGLERFAGSRLVVVTRGAVVAGPEDGPVDVVGASVWGLVRSAQAEHPDRFVLLDLDT 3441
Db       3160 QGWLGLERFAGSRLVVVTRGAVVAGPEDGPVDVVGAAVWGLVRSAQAEHPDRFVLLDLDT 3219

Qy       3442 --DTGTDLDTG--AGAGWGVDGGRVAAVVACGEPQLAVRGERLLAARLKRLESSGDVPAQ 3497
Db       3220 DLDSGADADAGNEAGMGSGLDGGRVAAVVACGEPQLAVRGERVLAARLTRLES------- 3272

Qy       3498 RSGDTRARRSDVPAQRSGGVPARRSVDVSGREVLPWLSGGSVLVTGGTGVLGAAVARHLA 3557
Db       3273 -----------------------PVDVSGREVLPWLSGGSVLVTGGTGVLGAAVARHLA 3308

Qy       3558 GVCGVRDLLLVSRRGPDAPGAEGLRAELAALGAEVRIVACDVGERREVVRLLEGVPAGCP 3617
Db       3309 GVCGVRDLLLVSRRGPDAPGAEGLRAELAALGAEVRIVACDVGERREVVRLLEGVPAGCP 3368

Qy       3618 LTGVVHAAGVLDDATIASLTPERLGTVFAAKVDAALLLDELTRGMELSAFVLFSSAAGIL 3677
Db       3369 LTGVVHAAGVLDDATIASLTPERLGTVFAAKVDAALLLDELTRGMELSAFVLFSSAAGIL 3428

Qy       3678 GSAGQGNYAAANAALDALAYRRRAAGLPGVSLAWGLWEEASGMTGHLAGTDHRRIIRSGL 3737
Db       3429 GSAGQGNYAAANAALDALAYRRRAAGLPGVSLAWGLWEEASGMTGHLAGTDHRRIIRSGL 3488

Qy       3738 HPMSTPDALALFDAALALDRPVLLPADLRPAPPLPPLLQDLLPATRRRTTRTTTTGGADN 3797
Db       3489 HPMSTPDALALFDAALALDRPVLLPADLRPAPPLPPLLQDLLPATRRRTTRTTTTGGADN 3548

Qy       3798 GAQLHARLAGQTHEQQHTTLLALVRSHIATVLGHTTPDTIPPDRAFRDLGFDSLTAVELR 3857
Db       3549 GAQLHARLAGQTHEQQHTTLLALVRSHIATVLGHNAPEMIPVDSAFRDLGFDSLTAVELR 3608

Qy       3858 NRLSRTTGLRLPTTLAFDHPNPTTLTHHLHTQL 3890
Db       3609 NRLGEATGLRLPTSLVFDQPNAATLARHLRREL 3641
```

RESULT 5
AAG65267
ID AAG65267 standard; Protein; 5532 AA.

XX
AC AAG65267;
XX
DT 04-DEC-2001 (first entry)
XX
DE Streptomyces avermitilis protein SEQ ID NO: 7.
XX
KW Avermectin aglycone synthase; AAS; avermectin derivative;
KW drug production; vetinary drug; pesticide.
XX
OS Streptomyces avermitilis.
XX
PN WO200162939-A1.
XX
PD 30-AUG-2001.
XX
PF 23-FEB-2001; 2001WO-JP01381.
XX
PR 24-FEB-2000; 2000JP-0047405.
XX
PA (KYOW) KYOWA HAKKO KOGYO KK.
PA (KITA) KITASATO INST.
XX
PI Endo H, Yamaguchi H, Kanda Y, Hashimoto S, Omura S, Ikeda H;
XX
DR WPI; 2001-582053/65.
DR N-PSDB; AAH79278.
XX
PT New modified avermectin aglycone synthase derived from Streptomyces
PT avermectilis used in production of 22,23-dihydroavermectin B1a used in
PT drugs and pesticides -
XX
PS Claim 4; Page 217-235; 257pp; Japanese.
XX
CC The present invention relates to the production of modified derivatives
CC of avermectin aglycone synthase (AAS) derived from Streptomyces
CC avermitilis. The activity of an acyl carrier protein (ACP),
CC beta-ketoacyl carrier protein synthase (KS), acyltransferase (AT),
CC beta-ketoacyl carrier protein reductase (KR), dehydratase (DH), enoyl
CC reductase (ER) and/or thioesterase (TE) domain may be reduced or
CC suppressed. The process can be used in the production of drugs, vetinary
CC drugs and pesticides. The present sequence is an S. avermitilis protein.
XX
SQ Sequence 5532 AA;

Query Match 62.3%; Score 12972; DB 22; Length 5532;
Best Local Similarity 71.5%; Pred. No. 0;
Matches 2640; Conservative 168; Mismatches 519; Indels 366; Gaps 47;

```
Qy        478 AADDPIAIIGMACRFPGGVRSPKDLWELAASGGDAIGPFPTDRGWPTEQRHAQDPTQPGT 537
Db         35 ASSEPIAIVGMACRFPGGVCSPEELWELVASGGDAIGEFPAGRGWDLEGLFDSDPDRSGT 94

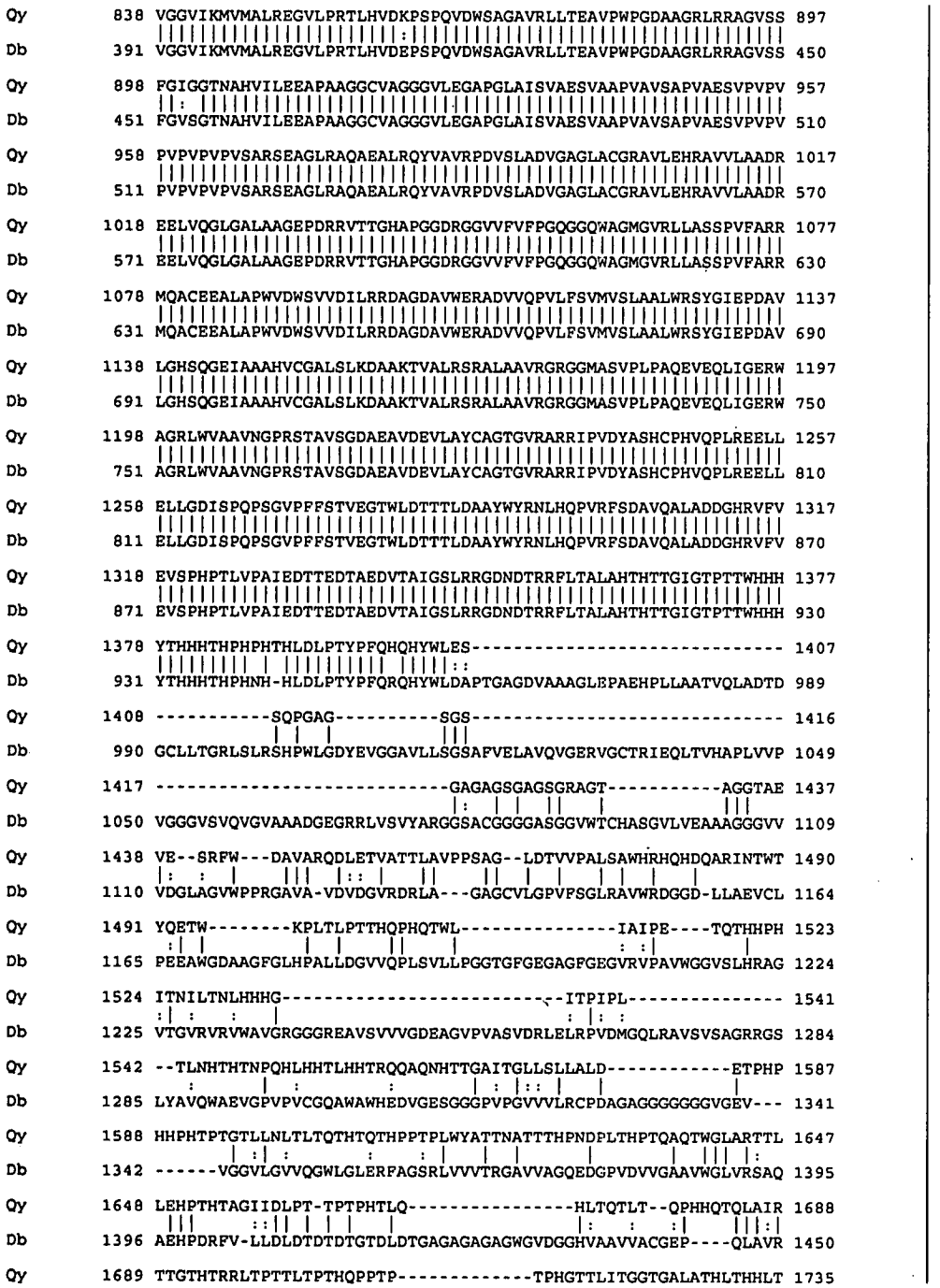
Qy        538 FYPQGGGFLHDAAHFDAGFFGISPREALAMDPQQRLLLETSWEAFERAGIDPLSVRGSRT 597
Db         95 SYARYGGFLYEAGEFDADFFGISPREALAMDPQQRLLLETSWEAFERAGIDPLSMRGSRT 154

Qy        598 GVFAGALSFDYGPRMDTASSEGAADVEGHILTGTTGSVLSGRIAYSFGLEGPAITVDTGC 657
Db        155 GVFAGVMYHDYGSRLGTI-PEG---FEGYIGNGSGGAVASGRVAYTLGLEGPAVSVDTAC 210

Qy        658 SASLVTLHLACQSLRSGECTLALAGGVSVMSTLGMFIEFSRQRGLSVDGRCKAYSAAADG 717
Db        211 SSSLVALHLACQSLRSGECTLALAGGVTVMSTPHLFVEFSRQRGLSVDGRCKSFAGGADG 270

Qy        718 TGWGEGVGMLLVERLSDAVRLGHRVLAVVRGSAVNQDGASNGLTAPNGPAQERVIRQALA 777
Db        271 TGMGEGVGMLLVERLSDAVRLGHRVLAVLRGSAVNQDGASNGLTAPNGPAQERVIRQALA 330

Qy        778 NAGLSVADVDVVEGHGTGTTLGDPIEAQALLATYGQRAGDRPLWLGSLKSNIGHTMAAAG 837
Db        331 NAGLSVADVDVVEGHGTGTTLGDPIEAQALLATYGQRAGNRPLWLGSVKSNIGHAQAAAG 390
```


Qy 838 VGGVIKMVMALREGVLPRTLHVDKPSPQVDWSAGAVRLLTEAVPWPGDAAGRLRRAGVSS 897
Db 391 VGGVIKMVMALREGVLPRTLHVDEPSPQVDWSAGAVRLLTEAVPWPGDAAGRLRRAGVSS 450
Qy 898 FGIGGTNAHVILEEAPAAGGCVAGGGVLEGAPGLAISVAESVAAPVAVSAPVAESVPVPV 957
Db 451 FGVSGTNAHVILEEAPAAGGCVAGGGVLEGAPGLAISVAESVAAPVAVSAPVAESVPVPV 510
Qy 958 PVPVPVPVSARSEAGLRAQAEALRQYVAVRPDVSLADVGAGLACGRAVLEHRAVVLAADR 1017
Db 511 PVPVPVPVSARSEAGLRAQAEALRQYVAVRPDVSLADVGAGLACGRAVLEHRAVVLAADR 570
Qy 1018 EELVQGLGALAAGEPDRRVTTGHAPGGDRGGVVFVFPGQGGQWAGMGVRLLASSPVFARR 1077
Db 571 EELVQGLGALAAGEPDRRVTTGHAPGGDRGGVVFVFPGQGGQWAGMGVRLLASSPVFARR 630
Qy 1078 MQACEEALAPWVDWSVVDILRRDAGDAVWERADVVQPVLFSVMVSLAALWRSYGIEPDAV 1137
Db 631 MQACEEALAPWVDWSVVDILRRDAGDAVWERADVVQPVLFSVMVSLAALWRSYGIEPDAV 690
Qy 1138 LGHSQGEIAAAHVCGALSLKDAAKTVALRSRALAAVRGRGGMASVPLPAQEVEQLIGERW 1197
Db 691 LGHSQGEIAAAHVCGALSLKDAAKTVALRSRALAAVRGRGGMASVPLPAQEVEQLIGERW 750
Qy 1198 AGRLWVAAVNGPRSTAVSGDAEAVDEVLAYCAGTGVRARRIPVDYASHCPHVQPLREELL 1257
Db 751 AGRLWVAAVNGPRSTAVSGDAEAVDEVLAYCAGTGVRARRIPVDYASHCPHVQPLREELL 810
Qy 1258 ELLGDISPQPSGVPFFSTVEGTWLDTTTLDAAYWYRNLHQPVRFSDAVQALADDGHRVFV 1317
Db 811 ELLGDISPQPSGVPFFSTVEGTWLDTTTLDAAYWYRNLHQPVRFSDAVQALADDGHRVFV 870
Qy 1318 EVSPHPTLVPAIEDTTEDTAEDVTAIGSLRRGDNDTRRFLTALAHTHTTGIGTPTTWHHH 1377
Db 871 EVSPHPTLVPAIEDTTEDTAEDVTAIGSLRRGDNDTRRFLTALAHTHTTGIGTPTTWHHH 930
Qy 1378 YTHHHTHPHPHTHLDLPTYPFQHQHYWLES---------- 1407
Db 931 YTHHHTHPHNH-HLDLPTYPFQRQHYWLDAPTGAGDVAAAGLEPAEHPLLAATVQLADTD 989
Qy 1408 -----------SQPGAG----------SGS---------- 1416
Db 990 GCLLTGRLSLRSHPWLGDYEVGGAVLLSGSAFVELAVQVGERVGCTRIEQLTVHAPLVVP 1049
Qy 1417 ----------GAGAGSGAGSGRAGT----------AGGTAE 1437
Db 1050 VGGGVSVQVGVAAADGEGRRLVSVYARGGSACGGGGASGGVWTCHASGVLVEAAAGGGVV 1109
Qy 1438 VE--SRFW---DAVARQDLETVATTLAVPPSAG--LDTVVPALSAWHRHQHDQARINTWT 1490
Db 1110 VDGLAGVWPPRGAVA-VDVDGVRDRLA---GAGCVLGPVFSGLRAVWRDGGD-LLAEVCL 1164
Qy 1491 YQETW--------KPLTLPTTHQPHQTWL----------IAIPE----TQTHHPH 1523
Db 1165 PEEAWGDAAGFGLHPALLDGVVQPLSVLLPGGTGPGEGAGFGEGVRVPAVWGGVSLHRAG 1224
Qy 1524 ITNILTNLHHHG----------ITIPIPL---------- 1541
Db 1225 VTGVRVRVWAVGRGGGREAVSVVVGDEAGVPVASVDRLELRPVDMGQLRAVSVSAGRRGS 1284
Qy 1542 --TLNHTHTNPQHLHHTLHHTRQQAQNHTTGAITGLLSLLALD----------ETPHP 1587
Db 1285 LYAVQWAEVGPVPVCGQAWAWHEDVGESGGGPVPGVVVLRCPDAGAGGGGGGGVGEV--- 1341
Qy 1588 HHPHTPTGTLLNLTLTQTHTQTHPPTPLWYATTNATTTHPNDPLTHPTQAQTWGLARTTL 1647
Db 1342 ------VGGVLGVVQGWLGLERFAGSRLVVVTRGAVVAGQEDGPVDVVGAAVWGLVRSAQ 1395
Qy 1648 LEHPTHTAGIIDLPT-TPTPHTLQ----------HLTQTLT--QPHHQTQLAIR 1688
Db 1396 AEHPDRFV-LLDLDTDTDTGTDLDTGAGAGAGAGWGVDGGHVAAVVACGEP----QLAVR 1450
Qy 1689 TTGTHTRRLTPTTLTPTHQPPTP----------TPHGTTLITGGTGALATHLTHHLT 1735
Db 1451 GERVLAARL--TRLESSVDVPAQRSGDVAGREVLPWLSGGSVLVTGGTGVLGAAVARHLA 1508
Qy 1736 THQPTQHLLLTSRTGPHTPHAQHLTTQLQQKGIHLTITTCDTSNPDQLQQLLNTIPPQHP 1795
Db 1509 GVCGVRDLLLVSRRGPDAPGAEGLRAELAALGAEVRIVACDVGERREVVRLLEGVPAGCP 1568
Qy 1796 LTTVIHTAGILDDATLTNLTPTQLNNVLRAKAHSAHLLHQLTQHTPLTAFVLYSSAAATF 1855
Db 1569 LTGVVHAAGVLDDATIASLTPERLGTVFAAKVDAALLLDELTRGMELSAFVLFSSAAGIL 1628
Qy 1856 GAPGQANYAAANAYLDALAHHRHTHHLPATSIAWGTW-QGNG----LADSDKARAYLDRR 1910
Db 1629 GSAGQGNYAAANAALDALAYRRRAAGLPGVSLAWGLWEEASGMTGHLAGTDHRR--IIRS 1686
Qy 1911 GFRPMSPELATAAVTQAIADTERPYVVIADIDWSKIEHTSQTSDLVSAARAREPAVQRPT 1970
Db 1687 GLHPMSTPDALALFDAALA-LDRPVLLPADL--------RPAPPLPPLLQDLLPATRRRT 1737
Qy 1971 P----------PAELHKTLAHQTSADQRAALLELVRDHVAAVLRHADPKAIAPDQSFRAL 2020
Db 1738 TRTTTTGGADNGAQLHARLAGQTHEQQHTTLLALVRSHIATVLGHTTPDTIPPDRAFRDL 1797
Qy 2021 GFDSLTAVEFRNLLIKATGLRLPVSLVFDHPTPAKLAVHLQNQLRG---TAAESAPSAAA 2077
Db 1798 GFDSLTAVELRNRLSRTTGLRLPTTLAFDHPNPTTLTHHLHTQLLGSDSTASIPAPRAAA 1857
Qy 2078 VTAEASVTEPIAIVGMACRFPGGVTSADDFWDLISSEQDAIGGFPTDRGWDLDTLYDPDP 2137
Db 1858 VPADQD--EPVAIIGMACRYPGGVTSAEELWELLASGRDTVGEFPTDRGWDLEALFDPEP 1915
Qy 2138 DHPGTCYTRNGGFLYDAGHFDAEFFGISPREALAMDPQQRLLLETAWETIEHAGINPHTL 2197
Db 1916 GRPGTSYTRCGSFLYDAGEFDAGFFGISPREALAMDPQQRLLLEASWEAMEQAGIDPTTV 1975
Qy 2198 HGTPTGVFTGTNGQDYALRVH-NAGQSTDGFALTGTAGSVISGRISYTFGFEGPAVSVDT 2256
Db 1976 RGSQTGVFAGLIPQAYGPRLHENAAADTEGYVLTGTSGSVASGRISYTFGFEGPAVSVDT 2035
Qy 2257 ACSSSLVALHLACQALRAGECSMALAGGVTVMSSPGAFVEFSRQRGLAADGHCKAFSAAA 2316
Db 2036 ACSSSLVALHLACQALRAGECSMALAGGVTVMSSPGAFVEFSRQRGLAADGHCKAFSAAA 2095
Qy 2317 DGTGWGEGVGMLLVERLSDAHRNGHRVLAVVRGSAVNQDGASNGLTAPNGPSQQRVIRQA 2376
Db 2096 DGTGWGEGVGMLLVERLSDARRNGHRVLAVVRGSAVNQDGASNGLTAPNGPSQQRVIRQA 2155
Qy 2377 LANAGLSAGDVDAVEAHGTGTTLGDPIEAQALLATYGQDRAGEGPLWLGSVKSNVGHTQA 2436
Db 2156 LANAGLSAGDVDAVEAHGTGTTLGDPIEAQALLATYGQDRAGEGPLWLGSVKSNVGHTQA 2215
Qy 2437 AAGVAGVIKMVMALRHGLLPRTLHVDEPSPHVDWSAGAVQLLTETVPWPGGEGRLRRAGV 2496
Db 2216 AAGVAGVIKMVMALRNGLLPRTLHVDEPSPHVDWSAGAVQLLTETVPWPGGEGRLRRAGV 2275
Qy 2497 SSFGVSGTNAHVILEEA---------PADDVPGGPPAGEGDAGSDDEAAAGSPGVWPWLV 2547
Db 2276 SSFGVSGTNAHVILEEAPAHNIPSDTPADDVPGGPPAGE-DAGSGEEAAAGSPGVWPWLV 2334
Qy 2548 SAKSQPALRAQAQALHAHLTDHPGLDLADVGYTLAHARAVFDHRATLIAADRDTFLQALQ 2607
Db 2335 SAKSQPALRAQAQALHAHLTDHPGLDLADVGYTLAHARAVFDHRATLIAADRDTFLQALQ 2394
Qy 2608 ALAAGEPHPAVIHSSAPGGTGTGEAAGKTAFICSGQGTQRPGMAHGLYHTHPVFAAALND 2667
Db 2395 ALAAGEPHPAVIHSSAPGGTGTGEAAGKTAFICSGQGTQRPGMAHGLYHTHPVFAAALND 2454
Qy 2668 ICTHLDPHLDHPLLPLLTQNDNDN-----EDAAALLQQTRYAQPALFAFQVALHRLLTDG 2722
Db 2455 ICTHLDPHLDHPLLPLLTQDPNTQDTTTLEEAAALLQQTPYAQPALFAFQVALHRLLTDG 2514
Qy 2723 YHITPHYYAGHSLGEITAAHLAGILTLTDATTLITQRATLMQTMPPGTMTTLHTTPHHIT 2782

```
Db       2515 YHITPHYYAGHSLGEITAAHLAGILTLTDATTLITQRATLMQTMPPGTMTTLHTTPHHIT 2574

Qy       2783 HHLTAHENDLAIAAINTPTSLVISGTPHTVQHITTLCQQQGIKTKTLPTNHAFHSPHTNP 2842
              ||:||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2575 HHITAHENDLAIAAINTPTSLVISGTPHTVQHITTLCQQQGIKTKTLPTNHAFHSPHTNP 2634

Qy       2843 ILNQLHQHTQTLTYHPPHTPLITANTPPDQLLTPHYWTQQARNTVDYATTTQTLHQHGVT 2902
              ||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db       2635 ILNQLHQHTQTLTYHPPHTPLITANTPPDQLLTPHYWTQQARNTVDIATTTQTLHQHGVT 2694

Qy       2903 TYIELGPDNTLTTLTHHNLPNPPTTTLTLTHPHHHPQTHLLTNLAKTTTTWHPHHYTHHD 2962
              |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db       2695 TYIELGPDNTLTTLTHHNLPNTPTTTLTLTHPHHHPQTHLLTNLAKTTTTWHPHHYTHHH 2754

Qy       2963 NQPHTHTHLDLPTYPFQHHHYWLESTQPGAGNVSAAGLDPTEHPLLGATLELATDGGALL 3022
              ||||||||||||||||||||||||||||||||||||||||||||||||||||   | ||
Db       2755 NQPHTHTHLDLPTYPFQHHHYWLESTQPGAGNVSAAGLDPTEHPLLGATLELAEGDGCLL 2814

Qy       3023 AGRLSLRSHPWLADHAVGGTVLLSGATFLELALHAGTYVGCDRVDELTLHAPLVVPVDGG 3082
              ||||||:||||| ||||| ||| |  | |||||||  |||| ||||||| |||:|  |
Db       2815 TGRLSLRTHPWLAGHAVGGVVLLPGTAFAELALHAGESVGCDHVDELTLHTPLVIPEVGD 2874

Qy       3083 VSVQVGVAAADGEGRRLVSVYARGGSACGGGGASGGVWTCHASGVLVEAAAGGVV-VDGL 3141
              |::|| :|| |  |||:::::||    ||  ||   || ||:||| |        : |
Db       2875 VTLQVAIAAPDESGRRMMTIHSRGEGGSGGADASASAWTRHAAGVLSPAKDDDTASYELL 2934

Qy       3142 AGVWPPRGAVAVDVDGVRDRLAGAGCVLGPVFSGLRAVWRDGGDLLAEVCLPEEAWGDAA 3201
              || ||| ||  ||::   |::| ||   |  | |||| || | |:| |  ||||  ||||
Db       2935 AGPWPPVGATPVDLNTAYDQMADAGFAYGLAFQGLRAAWRYGDDILVEARLPEEVSGDAA 2994

Qy       3202 GFGLHPALLDGVVQPLSVLLPGGTGFGEGAGFGEGVRVPAVWGGVSLHRAGVTGVRVRVS 3261
              :|||||||  :|        |||    || |  | || || |:      |   ||  :|
Db       2995 AYGLHPALLDAALQ--------GTGLLSVAGPGTPV-VPHVWNGLRFRTHGAVSVRACLS 3045

Qy       3262 AVGRGGGREAVSVVVGDEAGVPVASVDRLELRPVDMGQLRAVSVSAGRRGSLYAVQWAEV 3321
              :|  |   || | : |: ||||||||||||||||||||||||||||||||||||||||
Db       3046 TLGATGA--AVCVRITDDTGVPVASVDRLELRPVDMGQLRAVSVSAGRRGSLYAVQWAEV 3103

Qy       3322 GPVPVCGQAWAWHEDVGESGGGPVPGVVVLRCPDAGAGGGGGGGGGGGVGEVVGGVLGVV 3381
              ||||||||||||||||||||||||||||||||||    ||||||||||||||||||
Db       3104 GPVPVCGQAWAWHEDVGESGGGPVPGVVVLRCPDAGA----DGGGGGGVGEVVGGVLGVV 3159

Qy       3382 QGWLGLERFAGSRLVVVTRGAVVAGPEDGPVDVVGASVWGLVRSAQAEHPDRFVLLDLDT 3441
              ||||||||||||||||||||||||||||||||||||||:||||||||||||||||||||
Db       3160 QGWLGLERFAGSRLVVVTRGAVVAGPEDGPVDVVGAAVWGLVRSAQAEHPDRFVLLDLDT 3219

Qy       3442 --DTGTDLDTG--AGAGWGVDGGRVAAVVACGEPQLAVRGERLLAARLKRLESSGDVPAQ 3497
                |:| | | |  || | |:|||||||||||||||||||||:||||| ||||
Db       3220 DLDSGADADAGNEAGMGSGLDGGRVAAVVACGEPQLAVRGERVLAARLTRLES------- 3272

Qy       3498 RSGDTRARRSDVPAQRSGGVPARRSVDVSGREVLPWLSGGSVLVTGGTGVLGAAVARHLA 3557
                                  ||||||||||||||||||||||||||||||||||
Db       3273 ----------------------PVDVSGREVLPWLSGGSVLVTGGTGVLGAAVARHLA 3308

Qy       3558 GVCGVRDLLLVSRRGPDAPGAEGLRAELAALGAEVRIVACDVGERREVVRLLEGVPAGCP 3617
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3309 GVCGVRDLLLVSRRGPDAPGAEGLRAELAALGAEVRIVACDVGERREVVRLLEGVPAGCP 3368

Qy       3618 LTGVVHAAGVLDDATIASLTPERLGTVFAAKVDAALLLDELTRGMELSAFVLFSSAAGIL 3677
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3369 LTGVVHAAGVLDDATIASLTPERLGTVFAAKVDAALLLDELTRGMELSAFVLFSSAAGIL 3428

Qy       3678 GSAGQGNYAAANAALDALAYRRRAAGLPGVSLAWGLWEEASGMTGHLAGTDHRRIIRSGL 3737
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3429 GSAGQGNYAAANAALDALAYRRRAAGLPGVSLAWGLWEEASGMTGHLAGTDHRRIIRSGL 3488

Qy       3738 HPMSTPDALALFDAALALDRPVLLPADLRPAPPLPPLLQDLLPATRRRTTRTTTTGGADN 3797
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3489 HPMSTPDALALFDAALALDRPVLLPADLRPAPPLPPLLQDLLPATRRRTTRTTTTGGADN 3548

Qy       3798 GAQLHARLAGQTHEQQHTTLLALVRSHIATVLGHTTPDTIPPDRAFRDLGFDSLTAVELR 3857
              |||||||||||||||||||||||||||||||  |: || | ||||||||||||||||
Db       3549 GAQLHARLAGQTHEQQHTTLLALVRSHIATVLGHNAPEMIPVDSAFRDLGFDSLTAVELR 3608

Qy       3858 NRLSRTTGLRLPTTLAFDHPNPTTLTHHLHTQL 3890
                 |||   |||||||:| || ||  ||  ||  :|
Db       3609 NRLGEATGLRLPTSLVFDQPNAATLARHLRREL 3641

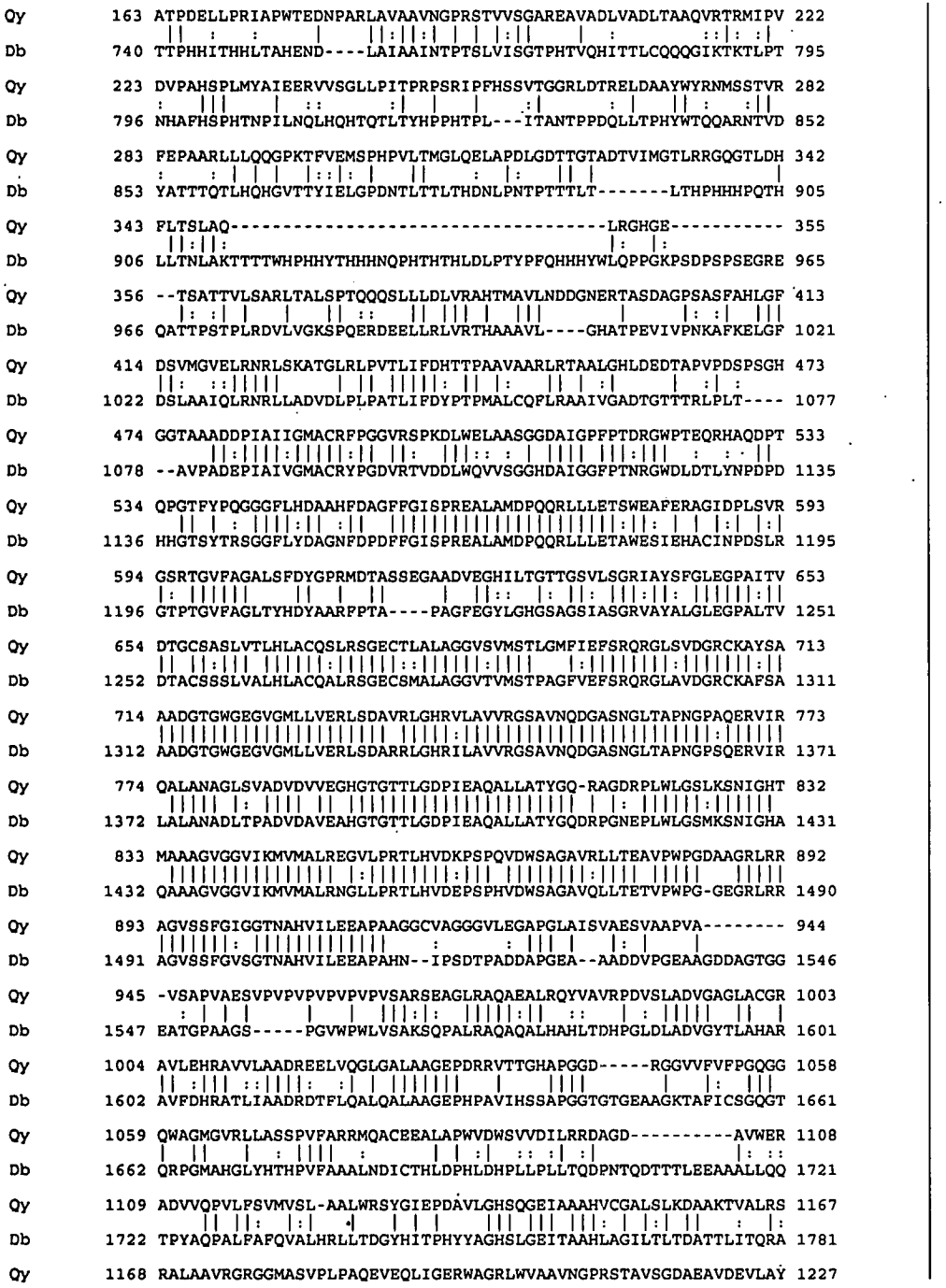

RESULT 6
AAB23750
ID   AAB23750 standard; Protein; 6239 AA.
XX
AC   AAB23750;
XX
DT   10-JAN-2001  (first entry)
XX
DE   S. avermitilis avermectin aglycon synthase protein SEQ ID NO:4.
XX
KW   Streptomyces avermitilis; avermectin aglycon synthase; biosynthesis;
KW   multifunctional enzyme; polyketide; avermectin; veterinary drug;
KW   agrochemical.
XX
OS   Streptomyces avermitilis.
XX
PN   WO200050605-A1.
XX
PD   31-AUG-2000.
XX
PF   23-FEB-2000; 2000WO-JP01041.
XX
PR   24-FEB-1999;   99JP-0046961.
XX
PA   (KITA ) KITASATO INST.
XX
PI   Omura S,  Ikeda H;
XX
DR   WPI; 2000-565458/52.
DR   N-PSDB; AAA92301.
XX
PT   Avermectin aglycone synthase DNA and proteins encoded by all or part of
PT   it for the production of avermectin and its derivatives for drug and
PT   agrochemical use  -
XX
PS   Claim 32; Page 223-254; 314pp; Japanese.
XX
CC   The present sequence represents an avermectin aglycon synthase protein.
CC   Also described are: (1) polypeptides encoded by all or part of the DNA;
CC   (2) expression vectors containing the DNA; (3) host cells transformed by
CC   the vectors; (4) preparation of the polypeptides by culture of the
CC   transformants; (5) preparation of avermectin aglycon or its derivatives
CC   by culture of transformed avermectin-producing microorganisms; and (6)
CC   oligonucleotides of 5-60 bases in length containing sense or antisense
CC   sequences from the avermectin aglycon synthase DNA. The enzymes are
CC   useful for the production of modified forms of avermectin and of the
CC   intermediates in its biosynthesis, for use as drugs, veterinary drugs
CC   and agrochemicals.
XX
SQ   Sequence   6239 AA;

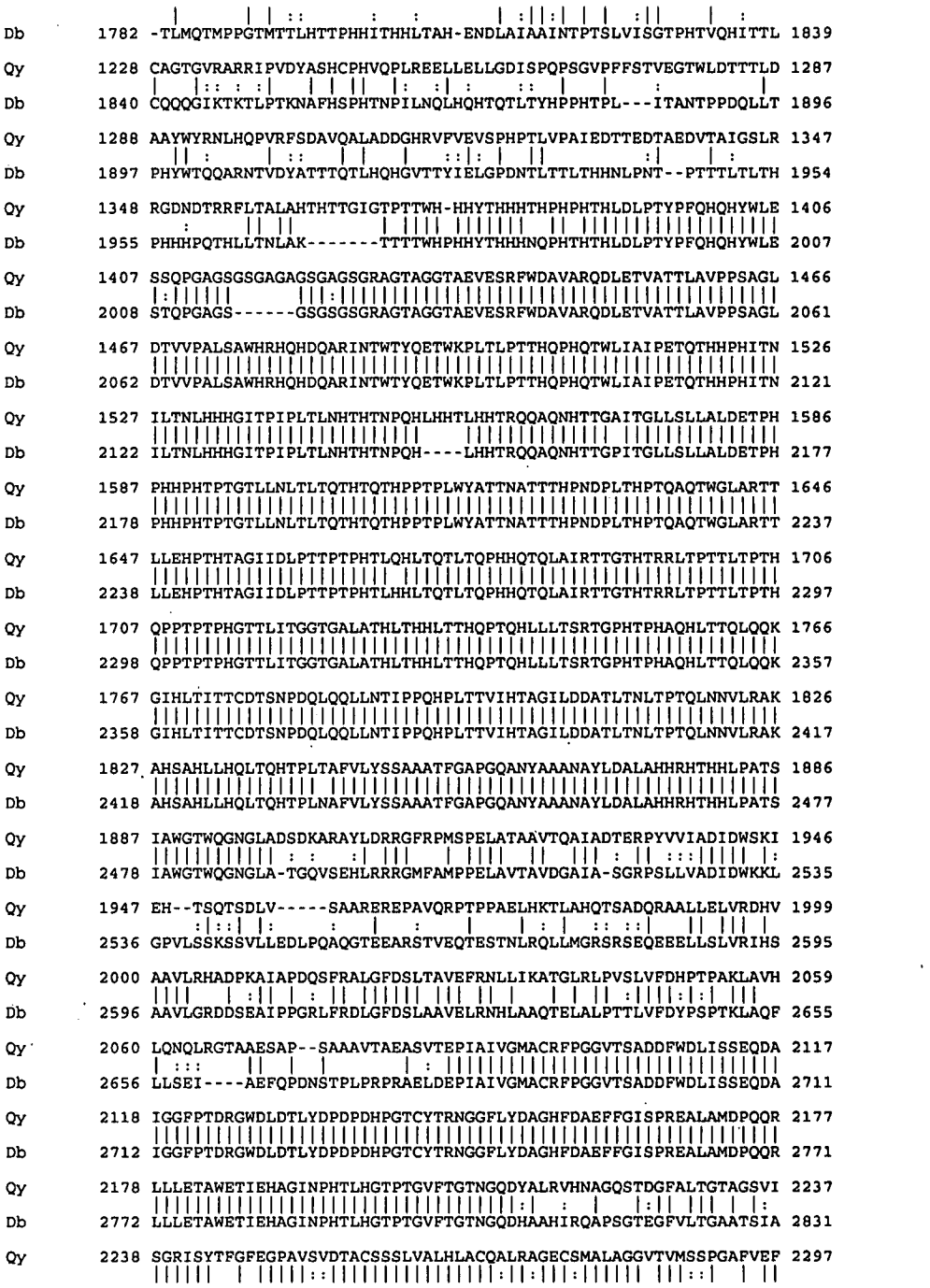
  Query Match             54.4%;  Score 11330.5;  DB 21;  Length 6239;
  Best Local Similarity   44.3%;  Pred. No. 0;
  Matches 2512;  Conservative  312;  Mismatches  961;  Indels 1883;  Gaps   61;

Qy         1 VQRMDGGEEPRPA-------AGEVLGVADEADGGVVFVFPGQGPQWPGMGRELLDASDVF 53
               :| :  | || ||       |   |   || |   |:  |:| | |||   |     ||
Db       565 LQALAAG-EPHPAVIHSSAPGGTGTG---EAAGKTAFICSGQGTQRPGMAHGLYHTHPVF 620

Qy        54 RESVRACEAAFAPYVDWSVEQVLRDSPD----------APGLDRVDVVQPTLFAVMISL- 102
             ::       |::|  :  :|   |:          |  | :    || |||  ::|
Db       621 AAALNDICTHLDPHLDHPLLPLLTQDPNTQDTTTLEEAAALLQQTRYAQPALFAFQVALH 680

Qy       103 AALWRSQGVEPCAVLGHSLGEIAAAHVSGGLSLADAARVVTLWSQAQTTLAGTGALVSVA 162
               |     : |    |||||||| |||::| |:| ||  ::|  :    |:   | : ::
Db       681 RLLTDGYHITPHYYAGHSLGEITAAHLAGILTLTDATTLITQRATLMQTMP-PGTMTTLH 739
```





```
Db       2832 SGRISYILGLEGPAVTLDTACSSSLVALHLACQSLRSGECTMALAGGATVMTTPITFTEF 2891

Qy       2298 SRQRGLAADGHCKAFSAAADGTGWGEGVGMLLVERLSDAHRNGHRVLAVVRGSAVNQDGA 2357
             :|||||| || |||||||||||||||||||||||||||| ||||||||||||||||||||
Db       2892 ARQRGLAPDGRCKAFSAAADGTGWGEGVGMLLVERLSDARRNGHRVLAVVRGSAVNQDGA 2951

Qy       2358 SNGLTAPNGPSQQRVIRQALANAGLSAGDVDAVEAHGTGTTLGDPIEAQALLATYGQDRA 2417
             |||||||||||||||||||||| |:  |||||||||||||||||||||:|||||||||
Db       2952 SNGLTAPNGPSQQRVIRQALANADLTPADVDAVEAHGTGTTLGDPIEAQAILATYGQDRP 3011

Qy       2418 GEGPLWLGSVKSNVGHTQAAAGVAGVIKMVMALRHGLLPRTLHVDEPSPHVDWSAGAVQL 2477
             | |||||||||||||||||||||||||||||||||   || ||| ||||||||||||||||
Db       3012 GNGPLWLGSVKSNVGHTQAAAGVAGVIKMVMALRHRTLPPTLHADEPSPHVDWSAGAVQL 3071

Qy       2478 LTETVPWPGGEGRLRRAGVSSFGVSGTNAHVILEEAPADDVPGGPPAGEGDAGSDDEAAA 2537
             |||||||||||| |||||||||||||||||||||||||||||| | |||| :||||
Db       3072 LTETVPWPGGEGRPRRAGVSSFGVSGTNAHVILEEAPADDVPGGPPADE-DAGSGEEAAA 3130

Qy       2538 GSPGVWPWLVSAKSQPALRAQAQALHAHLTDHPGLDLADVGYTLAHARAVFDHRATLIAA 2597
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3131 GSPGVWPWLVSAKSQPALRAQAQALHAHLTDHPGLDLADVGYTLAHARAVFDHRATLIAA 3190

Qy       2598 DRDTFLQALQALAAGEPHPAVIHSSAPGGTGTGEAAGKTAFICSGQGTQRPGMAHGLYHT 2657
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3191 DRDTFLQALQALAAGEPHPAVIHSSAPGGTGTGEAAGKTAFICSGQGTQRPGMAHGLYHT 3250

Qy       2658 HPVFAAALNDICTHLDPHLDHPLLPLLTQ--NDNDNEDAAALLQQTRYAQPALFAFQVAL 2715
             |||||||||||||||||||||||||||||  |||||||||||||| ||||||||||||
Db       3251 HPVFAAALNDICTHLDPHLDHPLLPLLTQNDNDNDNEDAAALLQQTPYAQPALFAFQVAL 3310

Qy       2716 HRLLTDGYHITPHYYAGHSLGEITAAHLAGILTLTDATTLITQRATLMQTMPPGTMTTLH 2775
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3311 HRLLTDGYHITPHYYAGHSLGEITAAHLAGILTLTDATTLITQRATLMQTMPPGTMTTLH 3370

Qy       2776 TTPHHITHHLTAHENDLAIAAINTPTSLVISGTPHTVQHITTLCQQQGIKTKTLPTNHAF 2835
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3371 TTPHHITHHLTAHENDLAIAAINTPTSLVISGTPHTVQHITTLCQQQGIKTKTLPTNHAF 3430

Qy       2836 HSPHTNPILNQLHQHTQTLTYHPPHTPLITANTPPDQLLTPHYWTQQARNTVDYATTTQT 2895
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3431 HSPHTNPILNQLHQHTQTLTYHPPHTPLITANTPPDQLLTPHYWTQQARNTVDYATTTQT 3490

Qy       2896 LHQHGVTTYIELGPDNTLTTLTHHNLPNPPTTTLTLTHPHHHPQTHLLTNLAKTTTTWHP 2955
             ||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db       3491 LHQHGVTTYIELGPDNTLTTLTHHNLPNTPTTTLTLTHPHHHPQTHLLTNLAKTTTTWHP 3550

Qy       2956 HHYTHHDNQPHTHTHLDLPTYPFQHHHYWLE---------------------------- 2986
             |||||| ||||||||||||||||||||||||
Db       3551 HHYTHHHNQPHTHTHLDLPTYPFQHHHYWLELPSAQTSPGQRRSRRSAPDTAESEFWDAV 3610

Qy       2987 ------------------------------------------------------------ 2986

Db       3611 NEEDLQSLAETLDIDASALDTVVPALSAWHRHQHDQARINTWTYQETWKPLTLPTTHQPH 3670

Qy       2987 ------------------------------------------------------------ 2986

Db       3671 QTWLIAIPETQTHHPHITNILTNLHHHGITPIPLTVNHTHTNPQHLHHTLHHTRQQAQNH 3730

Qy       2987 ------------------------------------------------------------ 2986

Db       3731 TTGPITGLLSLLALDETPHPHHPHTPTGTLLNLTLPQTHTQTHPPTPLWYATTNATTTHP 3790

Qy       2987 ----------------------------------------------------STQP-------- 2990
                                                                    |||
Db       3791 NDPLTHPTQAQTWGLARTTLLEHPTHTAGIIDLPTTPTPHTLHHLTQTLTQPHHQTQLAI 3850

Qy       2991 ------------------------------------------------------------ 2990

Db       3851 RTTGTHTRRLTPTTLTPTHQPPTPTPHGTTLITGGTGALATHLTHHLTTHQPTQHLLLTS 3910

Qy       2991 ------------------------------------------------------------ 2990

Db       3911 RTGPHTPHAQHLTTQLQQKGIHLTITTCDTSNPDQLQQLLNTIPPQHPLTTVIHTAGVNL 3970
```

```
Qy       2991 ----------------------------------------------------GAGN------- 2994
                                                                    |:||
Db       3971 FAPVSETDAESFSSVTAAKATGAAILHELLLDHETLEHFILFSSGAGAWGSGNQCAYSAA 4030

Qy       2995 ------------------------------------------------------------ 2994

Db       4031 NAYLDALATHRQTHGLPGASIAWGPWAGKGMSAGDAAHGYLEKRGILPMEPRMALAAFHR 4090

Qy       2995 ------------------------------------------------------------ 2994

Db       4091 ARAQRPNSNLIIADIDWERFVPAFTARRHSPLIEDIPEVRQAAQELEAAASTAKTTTAQP 4150

Qy       2995 ------------------------------------------------------------ 2994

Db       4151 IATSLRERLARLTSSKQNQVLLGLIRTGICTVLGLRNPEGIEDQRAFRDLGFDSLTSAQF 4210

Qy       2995 ------------------------------------------------------------ 2994

Db       4211 SKELAKETGLPLPPSLVFDYPTPQECAAHLRTQLVDLDDEEDAALSNALPQVAHRRTVED 4270

Qy       2995 ------------------------------------------------------------ 2994

Db       4271 EPIAIIGMACRFPGGVRSADDLWELLASGKDAIGVFPTDRGWDLDTLYDPDPDHPGTCYT 4330

Qy       2995 ------------------------------------------------------------ 2994

Db       4331 RNGGFLYGAGHFDAEFFGISPREALAMDPQQRLLLETAWETIEHAGINPHTLHGTPTGVF 4390

Qy       2995 ------------------------------------------------------------ 2994

Db       4391 AGINAQDHAAHIRQSRDVETIEGYALTGSSGSVASGRVAYTLGLEGPAVSVDTACSSSLV 4450

Qy       2995 ------------------------------------------------------------ 2994

Db       4451 ALHWAAQALRAGECSMALAGGVTVMSSPGTFVEFSRQRGLAADGRCKAYSAAADGTGWAE 4510

Qy       2995 ------------------------------------------------------------ 2994

Db       4511 GVGMLLVERLSDARRNGHRVLAVVRGSAVNQDGASNGLTAPNGPSQQRVIRQALANAGLT 4570

Qy       2995 ------------------------------------------------------------ 2994

Db       4571 PADVDAVEGHGTGTTLGDPIEAQALLAAYGQHRPHHRPLWLGSLKSNIGHAQAAAGVGGV 4630

Qy       2995 ------------------------------------------------------------ 2994

Db       4631 IKMVMALRNGLLPQTLHVDEPTPQVDWSTGAVQLLTQPVPWPADPAGRPRHAGVSSFGVS 4690

Qy       2995 ------------------------------------------------------------ 2994

Db       4691 GTNAHIILEEAPTPQDSDTDDEPPANAPALPHPLPLPVPVSARSEAGLRAQAQALRQYVA 4750

Qy       2995 ------------------------------------------------------------ 2994

Db       4751 ARPDMSPADIGAGLARGRAVLEHRAVILAADREELAQALTALAAGEPHPHITTGHTRGGD 4810

Qy       2995 ------------------------------------------------------------ 2994

Db       4811 RGGVVFVFPGQGGQWAGMGLTLLTSSPVFAEHIDACEKALTPWVPWSLTDILHRDPDDPA 4870

Qy       2995 ------------------------------------------------------------ 2994

Db       4871 WQQADVVQPVLFSIMVSLAALWRSYGIEPDAVLGHSQGEIAAAHICGALSLKDAAKTVAL 4930

Qy       2995 ------------------------------------------------------------ 2994

Db       4931 RSRALAAVRGRGAMASLPLPAQDVQQLISERWEGQLWVAALNGPHSTTVSGDTKAVDEVL 4990

Qy       2995 ------------------------------------------------------------ 2994

Db       4991 AHCTDTGLRAKRIPVDYASHCPHVQPLHDELLHLLGDITPQPSTVPFFSTVEGTWLDTTT 5050
```

```
Qy       2995 ------------------------------------------------------------ 2994

Db       5051 LDAAYWYRNLHQPVRFSHAIQTLTDDGHRAFIEISPHPTLVPAIEDTTENTTENITATGS 5110

Qy       2995 ------------------------------------------------------------ 2994

Db       5111 LRRGDNDTHRFLTALAHTHTTGIGTPTTWHHHYTQTHPHPNPHTHLDLPTYPFQHQHYWL 5170

Qy       2995 --------VSAAGLDPTEHPLLGATLELATDGGALLAGRLSLRSHPWLADHAVGGTVLLS 3046
                      ::  || || |||| ||| || :   || ||||||:|||| || | | |||
Db       5171 QPPTTTTDLTTTGLTPTHHPLLTATLTLADNNTQLLTGRLSLRTHPWLTDHTVAGMVLLP 5230

Qy       3047 GATFLELALHAGTYVGCDRVDELTLHAPLVVPVDGGVSVQVGVAAADGEGRRLVSVYARG 3106
              |   ||||| ||  | | ||:|||||||||:|    |::|| | |||  | | :::::
Db       5231 GTALLELALQAGERVDCPRVEELTLHAPLVIPHTEDVTLQVTVRAADESGHRALAIHSYS 5290

Qy       3107 GSACGGGGASGGVWTCHASGVLVEAAAGGVVVDG-----LAGVWPPRGAVAVDVDGVRDR 3161
              |:|     ::   || ||:|:|   |       |       | | ||| ||  ::: |  |
Db       5291 GTA----SSADREWTRHATGLLTHHADTDHRADTHTDACLGGSWPPPGAQPIELGDVYGR 5346

Qy       3162 LAG-AGCVLGPVFSGLRAVWRDGGDLLAEVCLPEEAWGD--AAGFGLHPALLDGVVQPLS 3218
              :|  :    |||| || | || | |:|||| ||||  |  || ||:||||||  :   :
Db       5347 MAADSDIAYGPVFQGLHAAWRFGDDVLAEVRLPEEALRDAPAAAFGVHPALLDAALHATA 5406

Qy       3219 VLLPGGTGFGEGAG-------FGEGVRVPAVWGGVSLHRAGVTGVRVRVSAVGRGGGREA 3271
              :    | |  |              |:|  | ||||| || : :|||:|   : |   |
Db       5407 LTPQNGDGSTENVAQESMPDRAAHQARLPFSWSGVSLHTAGSSVLRVRLSRSPQHG--NA 5464

Qy       3272 VSVVVGDEAGVPVASVDRLELRPVDMGQLRAVSVSAGRRGSLYAVQWAEV-----GPVPV 3326
              |::   || | || ::: | ||||   :||| :        ||: : |  |       | |
Db       5465 VALTAADEDGRPVVTIESLALRPVSTEELRAAADRTPEHESLFRLDWVSVPVPANAPSPT 5524

Qy       3327 CGQAWA-------------WHEDV-----------GESGGGPVPGVVVL--------RCP 3354
                : ||              || |                | | |||:|:          |
Db       5525 ADRPWAVIGAGLPHLPGLTEHEHVTAYDEPADLLLALDRGAPPPGVLVVGGVAHTEAREY 5584

Qy       3355 DAGAGGGGGGGGGGGVGEVV--------------GGVLGVVQGWLGLERFAGSRLVVVTR 3400
              | | |  |          :||                :|  :| ||| || | |||:|:|
Db       5585 SAEAPGERGTEACEARPDVVHVGVVHTAAVHAAAAQMLARLQAWLGDERLADSRLLVLTC 5644

Qy       3401 GAVVAGPEDGPVDVVGASVWGLVRSAQAEHPDRFVLLDLDTDTGTDLDTGAGAGWGVDGG 3460
              |||      |   |: ||:|||||||||:|||||  |||   : ||:           : |
Db       5645 GAVARASGDDATDLPGAAVWGLVRSAQSEHPDRITLLDF--ERGTE----------AEPG 5692

Qy       3461 RVAAVVACGEPQLAVRGERLLAARLKRLESSGDVPAQRSGDTRARRSDVPAQRSGGVPA- 3519
              ::|  : ||| |||||   |   || |         | |  |         |||  :  ||:
Db       5693 QLATALNCGERQLAVRPGGLFTPRLVR--------APRVADA------VPAVPAVAVPSA 5738

Qy       3520 -RRSVDVSGREVLPWLSGGSVLVTGGTGVLGAAVARHLAGVCGVRDLLLVSRRGPDAPGA 3578
               :|  :|    |:| ||:||:|||||||  |||||    ||| |||  |||||| ||
Db       5739 GHAAVPAAG----PFLPGGTVLITGGTGVLGRLVARHLVEAHGVRHLLLAGRRGPDAEGA 5794

Qy       3579 EGLRAELAALGAEVRIVACDVGERREVVRLLEGVPAGCPLTGVVHAAGVLDDATIASLTP 3638
                |||||  ||| | :|||| :|:::  ||  :|   |||||||:||:|||  | ||:|
Db       5795 PELRAELGGLGATVEVVACDAADRQQLADLLTRIPDDRPLTGVVHSAGILDDGVITSLSP 5854

Qy       3639 ERLGTVFAAKVDAALLLDELTRGMELSAFVLFSSAAGILGSAGQGNYAAANAALDALAYR 3698
              |||| |  || |||||||||||| ||||||:||||: ::|| |||||||||| || ||:|
Db       5855 ERLGAVLRAKADAALLLDELTRGAELSAFVMFSSASAVVGSPGQGNYAAANAVLDFLAHR 5914

Qy       3699 RRAAGLPGVSLAWGLWEEASGMTGHLAGTDHRRIIRSGLHPMSTPDALALFDAALALDRP 3758
              ||| ||| |||||||||| :|||||   || || |:|: |: | :|||||||||    |
Db       5915 RRAEGLPAVSLAWGLWEEGTGMTGHLDVDDHARISRAGMRPLPTAEALALFDAALADGEP 5974

Qy       3759 VLLPADL--------RPAPPLPPLLQDLLPATRRRTTRTTTTGG---ADNGAQLHARLAG 3807
               |:|| |          : |:||||| ||   | |:       |   ||  |    |||
Db       5975 FLMPARLDLTAVRSGAASAPVPPLLQGLLQLPRSRSAAAAPGHGAPAADEAAAWRERLAR 6034

Qy       3808 QTHEQQHTTLLALVRSHIATVLGHTTPDTIPPDRAFRDLGFDSLTAVELRNRLSRTTGLR 3867
              |:  ::   || |||||:| ||||:  | |   ||||:|||||||||||||:  ||||
Db       6035 QSAGERRQALLRLVRSHVAAVLGHSGADGIDASRAFRELGFDSLTAVELRNRLTAATGLR 6094

Qy       3868 LPTTLAFDHPNPTTLTHHLHTQLQPQPD 3895

                | |||||| | |  |  ||  :| | :
Db       6095 LRATLAFDFPTPAALAEHLGERLLPDQE 6122
```

RESULT 7
AAG65265
ID AAG65265 standard; Protein; 6239 AA.
XX
AC AAG65265;
XX
DT 04-DEC-2001 (first entry)
XX
DE Streptomyces avermitilis protein SEQ ID NO: 5.
XX
KW Avermectin aglycone synthase; AAS; avermectin derivative;
KW drug production; vetinary drug; pesticide.
XX
OS Streptomyces avermitilis.
XX
PN WO200162939-A1.
XX
PD 30-AUG-2001.
XX
PF 23-FEB-2001; 2001WO-JP01381.
XX
PR 24-FEB-2000; 2000JP-0047405.
XX
PA (KYOW) KYOWA HAKKO KOGYO KK.
PA (KITA) KITASATO INST.
XX
PI Endo H, Yamaguchi H, Kanda Y, Hashimoto S, Omura S, Ikeda H;
XX
DR WPI; 2001-582053/65.
DR N-PSDB; AAH79277.
XX
PT New modified avermectin aglycone synthase derived from Streptomyces
PT avermectilis used in production of 22,23-dihydroavermectin B1a used in
PT drugs and pesticides -
XX
PS Claim 4; Page 180-201; 257pp; Japanese.
XX
CC The present invention relates to the production of modified derivatives
CC of avermectin aglycone synthase (AAS) derived from Streptomyces
CC avermitilis. The activity of an acyl carrier protein (ACP),
CC beta-ketoacyl carrier protein synthase (KS), acyltransferase (AT),
CC beta-ketoacyl carrier protein reductase (KR), dehydratase (DH), enoyl
CC reductase (ER) and/or thioesterase (TE) domain may be reduced or
CC suppressed. The process can be used in the production of drugs, vetinary
CC drugs and pesticides. The present sequence is an S. avermitilis protein.
XX
SQ Sequence 6239 AA;

```
  Query Match             54.4%;  Score 11330.5;  DB 22;  Length 6239;
  Best Local Similarity   44.3%;  Pred. No. 0;
  Matches 2512;  Conservative  312;  Mismatches  961;  Indels 1883;  Gaps   61;

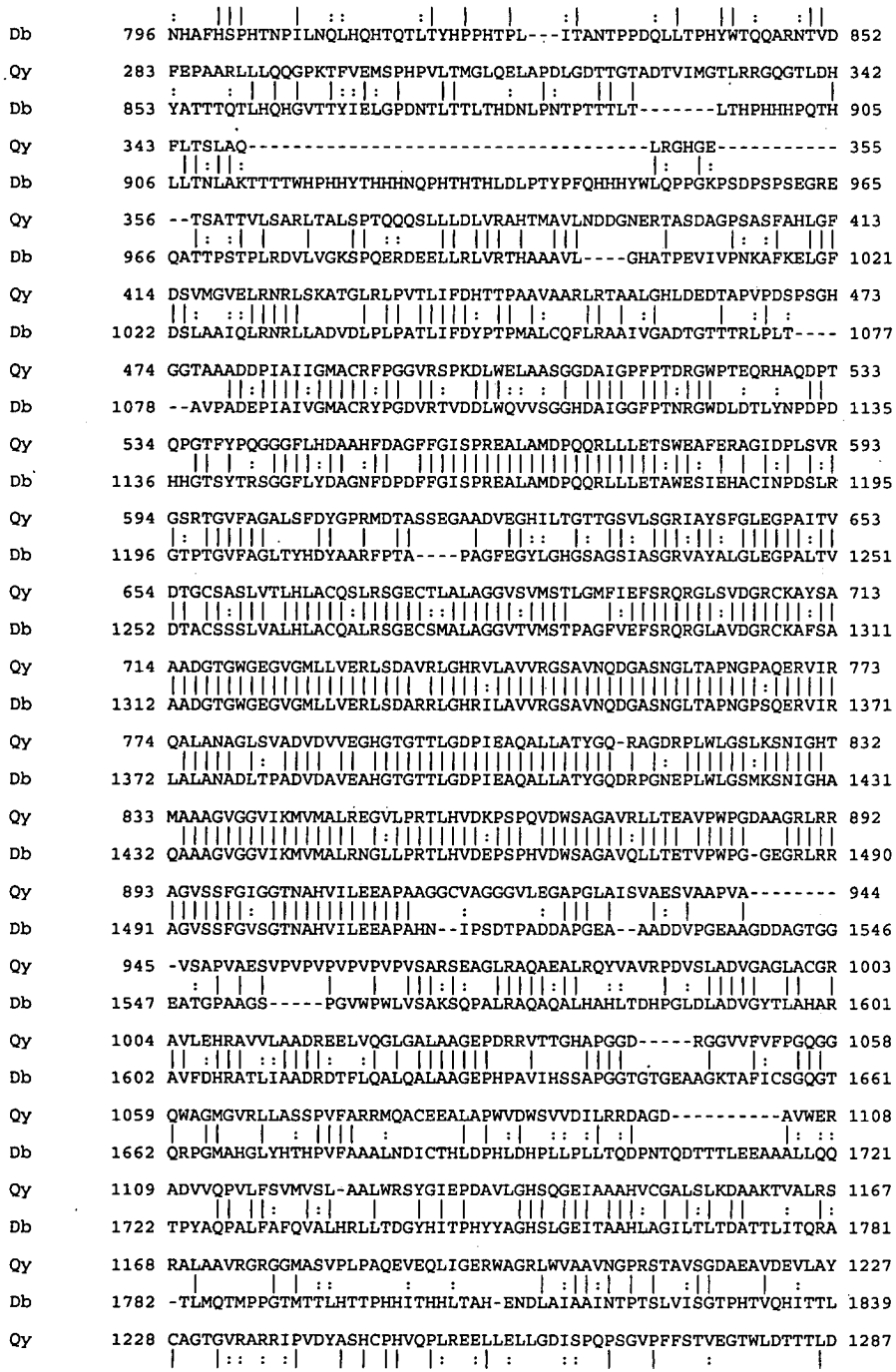
Qy          1 VQRMDGGEEPRPA-------AGEVLGVADEADGGVVFVFPGQGPQWPGMGRELLDASDVF 53
                :| :  | || ||        |   |   || |   |:  ||| | |||   |     ||
Db        565 LQALAAG-EPHPAVIHSSAPGGTGTG---EAAGKTAFICSGQGTQRPGMAHGLYHTHPVF 620

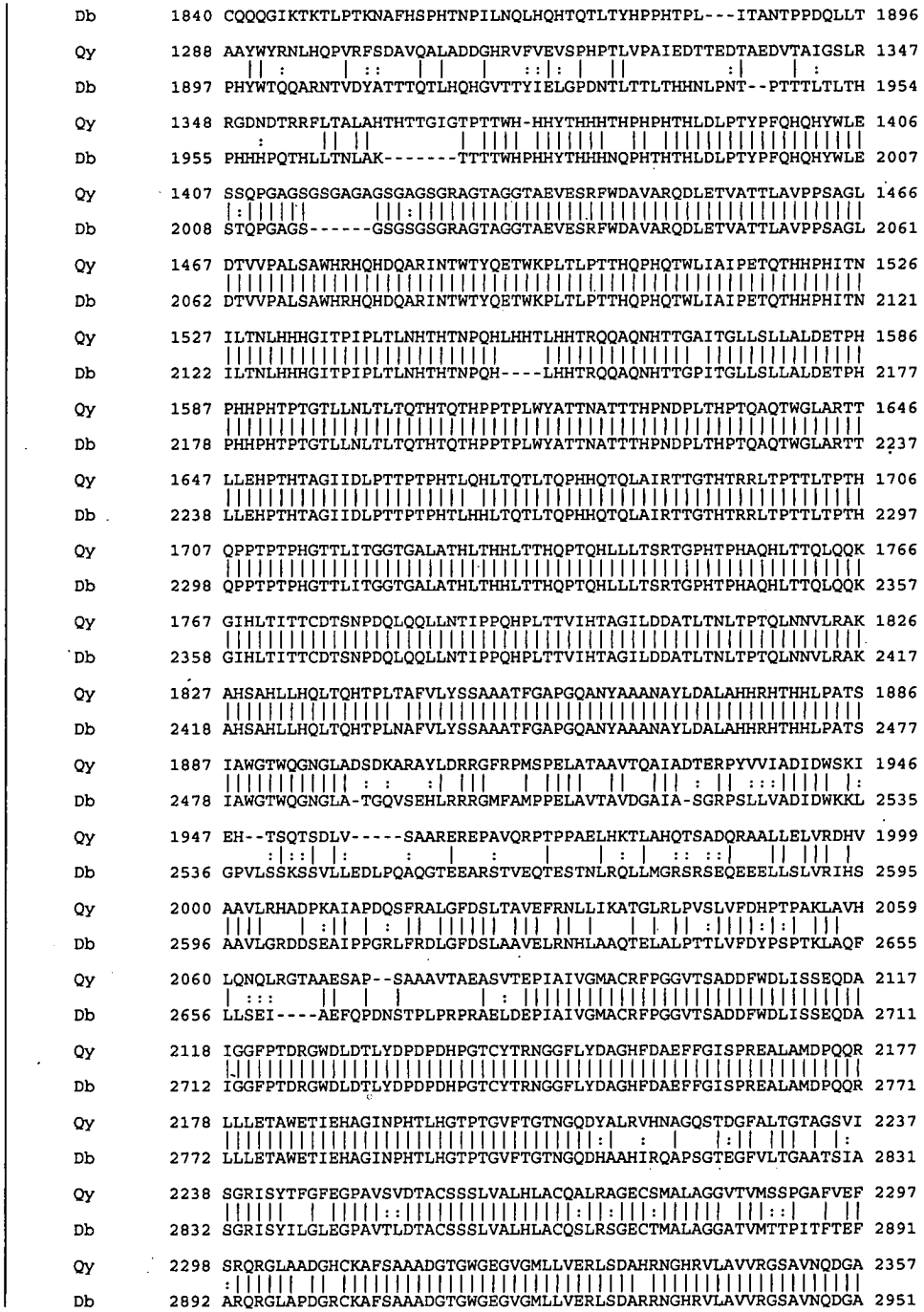
Qy         54 RESVRACEAAFAPYVDWSVEQVLRDSPD----------APGLDRVDVVQPTLFAVMISL- 102
                ::         |::|  :  :|   |:          |  | :    || |||  ::|
Db        621 AAALNDICTHLDPHLDHPLLPLLTQDPNTQDTTTLEEAAALLQQTRYAQPALFAFQVALH 680

Qy        103 AALWRSQGVEPCAVLGHSLGEIAAAHVSGGLSLADAARVVTLWSQAQTTLAGTGALVSVA 162
                |     : |    ||||||| |||::| |:| ||  ::|  :    |:   | : ::
Db        681 RLLTDGYHITPHYYAGHSLGEITAAHLAGILTLTDATTLITQRATLMQTMP-PGTMTTLH 739

Qy        163 ATPDELLPRIAPWTEDNPARLAVAAVNGPRSTVVSGAREAVADLVADLTAAQVRTRMIPV 222
                ||  :   :     |    ||:||:| | | |:||    |  :         ::|: :|
Db        740 TTPHHITHHLTAHEND----LAIAAINTPTSLVISGTPHTVQHITTLCQQQGIKTKTLPT 795

Qy        223 DVPAHSPLMYAIEERVVSGLLPITPRPSRIPFHSSVTGGRLDTRELDAAYWYRNMSSTVR 282
```


Db 796 NHAFHSPHTNPILNQLHQHTQTLTYHPPHTPL---ITANTPPDQLLTPHYWTQQARNTVD 852
Qy 283 FEPAARLLLQQGPKTFVEMSPHPVLTMGLQELAPDLGDTTGTADTVIMGTLRRGQGTLDH 342
Db 853 YATTTQTLHQHGVTTYIELGPDNTLTTLTHDNLPNTPTTTLT-------LTHPHHHPQTH 905
Qy 343 FLTSLAQ----------LRGHGE----------- 355
Db 906 LLTNLAKTTTTWHPHHYTHHHNQPHTHTHLDLPTYPFQHHHYWLQPPGKPSDPSPSEGRE 965
Qy 356 --TSATTVLSARLTALSPTQQQSLLLDLVRAHTMAVLNDDGNERTASDAGPSASFAHLGF 413
Db 966 QATTPSTPLRDVLVGKSPQERDEELLRLVRTHAAAVL----GHATPEVIVPNKAFKELGF 1021
Qy 414 DSVMGVELRNRLSKATGLRLPVTLIFDHTTPAAVAARLRTAALGHLDEDTAPVPDSPSGH 473
Db 1022 DSLAAIQLRNRLLADVDLPLPATLIFDYPTPMALCQFLRAAIVGADTGTTTRLPLT---- 1077
Qy 474 GGTAAADDPIAIIGMACRFPGGVRSPKDLWELAASGGDAIGPFPTDRGWPTEQRHAQDPT 533
Db 1078 --AVPADEPIAIVGMACRYPGDVRTVDDLWQVVSGGHDAIGGFPTNRGWDLDTLYNPDPD 1135
Qy 534 QPGTFYPQGGGFLHDAAHFDAGFFGISPREALAMDPQQRLLLETSWEAFERAGIDPLSVR 593
Db 1136 HHGTSYTRSGGFLYDAGNFDPDFFGISPREALAMDPQQRLLLETAWESIEHACINPDSLR 1195
Qy 594 GSRTGVFAGALSFDYGPRMDTASSEGAADVEGHILTGTTGSVLSGRIAYSFGLEGPAITV 653
Db 1196 GTPTGVFAGLTYHDYAARFPTA----PAGFEGYLGHGSAGSIASGRVAYALGLEGPALTV 1251
Qy 654 DTGCSASLVTLHLACQSLRSGECTLALAGGVSVMSTLGMFIEFSRQRGLSVDGRCKAYSA 713
Db 1252 DTACSSSLVALHLACQALRSGECSMALAGGVTVMSTPAGFVEFSRQRGLAVDGRCKAFSA 1311
Qy 714 AADGTGWGEGVGMLLVERLSDAVRLGHRVLAVVRGSAVNQDGASNGLTAPNGPAQERVIR 773
Db 1312 AADGTGWGEGVGMLLVERLSDARRLGHRILAVVRGSAVNQDGASNGLTAPNGPSQERVIR 1371
Qy 774 QALANAGLSVADVDVVEGHGTGTTLGDPIEAQALLATYGQ-RAGDRPLWLGSLKSNIGHT 832
Db 1372 LALANADLTPADVDAVEAHGTGTTLGDPIEAQALLATYGQDRPGNEPLWLGSMKSNIGHA 1431
Qy 833 MAAAGVGGVIKMVMALREGVLPRTLHVDKPSPQVDWSAGAVRLLTEAVPWPGDAAGRLRR 892
Db 1432 QAAAGVGGVIKMVMALRNGLLPRTLHVDEPSPHVDWSAGAVQLLTETVPWPG-GEGRLRR 1490
Qy 893 AGVSSFGIGGTNAHVILEEAPAAGGCVAGGGVLEGAPGLAISVAESVAAPVA-------- 944
Db 1491 AGVSSFGVSGTNAHVILEEAPAHN--IPSDTPADDAPGEA--AADDVPGEAAGDDAGTGG 1546
Qy 945 -VSAPVAESVPVPVPVPVPVPVSARSEAGLRAQAEALRQYVAVRPDVSLADVGAGLACGR 1003
Db 1547 EATGPAAGS-----PGVWPWLVSAKSQPALRAQAQALHAHLTDHPGLDLADVGYTLAHAR 1601
Qy 1004 AVLEHRAVVLAADREELVQGLGALAAGEPDRRVTTGHAPGGD-----RGGVVFVFPGQGG 1058
Db 1602 AVFDHRATLIAADRDTFLQALQALAAGEPHPAVIHSSAPGGTGTGEAAGKTAFICSGQGT 1661
Qy 1059 QWAGMGVRLLASSPVFARRMQACEEALAPWVDWSVVDILRRDAGD----------AVWER 1108
Db 1662 QRPGMAHGLYHTHPVFAAALNDICTHLDPHLDHPLLPLLTQDPNTQDTTTLEEAAALLQQ 1721
Qy 1109 ADVVQPVLFSVMVSL-AALWRSYGIEPDAVLGHSQGEIAAAHVCGALSLKDAAKTVALRS 1167
Db 1722 TPYAQPALFAFQVALHRLLTDGYHITPHYYAGHSLGEITAAHLAGILTLTDATTLITQRA 1781
Qy 1168 RALAAVRGRGGMASVPLPAQEVEQLIGERWAGRLWVAAVNGPRSTAVSGDAEAVDEVLAY 1227
Db 1782 -TLMQTMPPGTMTTLHTTPHHITHHLTAH-ENDLAIAAAINTPTSLVISGTPHTVQHITTL 1839
Qy 1228 CAGTGVRARRIPVDYASHCPHVQPLREELLELLGDISPQPSGVPFFSTVEGTWLDTTTLD 1287


Db 1840 CQQQGIKTKTLPTKNAFHSPHTNPILNQLHQHTQTLTYHPPHTPL---ITANTPPDQLLT 1896
Qy 1288 AAYWYRNLHQPVRFSDAVQALADDGHRVFVEVSPHPTLVPAIEDTTEDTAEDVTAIGSLR 1347
Db 1897 PHYWTQQARNTVDYATTTQTLHQHGVTTYIELGPDNTLTTLTHHNLPNT--PTTTLTLTH 1954
Qy 1348 RGDNDTRRFLTALAHTHTTGIGTPTTWH-HHYTHHHTHPHPHTHLDLPTYPFQHQHYWLE 1406
Db 1955 PHHHPQTHLLTNLAK-------TTTTWHPHHYTHHHNQPHTHTHLDLPTYPFQHQHYWLE 2007
Qy 1407 SSQPGAGSGSGAGAGSGAGSGRAGTAGGTAEVESRFWDAVARQDLETVATTLAVPPSAGL 1466
Db 2008 STQPGAGS------GSGSGSGRAGTAGGTAEVESRFWDAVARQDLETVATTLAVPPSAGL 2061
Qy 1467 DTVVPALSAWHRHQHDQARINTWTYQETWKPLTLPTTHQPHQTWLIAIPETQTHHPHITN 1526
Db 2062 DTVVPALSAWHRHQHDQARINTWTYQETWKPLTLPTTHQPHQTWLIAIPETQTHHPHITN 2121
Qy 1527 ILTNLHHHGITPIPLTLNHTHTNPQHLHHTLHHTRQQAQNHTTGAITGLLSLLALDETPH 1586
Db 2122 ILTNLHHHGITPIPLTLNHTHTNPQH----LHHTRQQAQNHTTGPITGLLSLLALDETPH 2177
Qy 1587 PHHPHTPTGTLLNLTLTQTHTQTHPPTPLWYATTNATTTHPNDPLTHPTQAQTWGLARTT 1646
Db 2178 PHHPHTPTGTLLNLTLTQTHTQTHPPTPLWYATTNATTTHPNDPLTHPTQAQTWGLARTT 2237
Qy 1647 LLEHPTHTAGIIDLPTTPTPHTLQHLTQTLTQPHHQTQLAIRTTGTHTRRLTPTTLTPTH 1706
Db 2238 LLEHPTHTAGIIDLPTTPTPHTLHHLTQTLTQPHHQTQLAIRTTGTHTRRLTPTTLTPTH 2297
Qy 1707 QPPTPTPHGTTLITGGTGALATHLTHHLTTHQPTQHLLLTSRTGPHTPHAQHLTTQLQQK 1766
Db 2298 QPPTPTPHGTTLITGGTGALATHLTHHLTTHQPTQHLLLTSRTGPHTPHAQHLTTQLQQK 2357
Qy 1767 GIHLTITTCDTSNPDQLQQLLNTIPPQHPLTTVIHTAGILDDATLTNLTPTQLNNVLRAK 1826
Db 2358 GIHLTITTCDTSNPDQLQQLLNTIPPQHPLTTVIHTAGILDDATLTNLTPTQLNNVLRAK 2417
Qy 1827 AHSAHLLHQLTQHTPLTAFVLYSSAAATFGAPGQANYAAANAYLDALAHHRHTHHLPATS 1886
Db 2418 AHSAHLLHQLTQHTPLNAFVLYSSAAATFGAPGQANYAAANAYLDALAHHRHTHHLPATS 2477
Qy 1887 IAWGTWQGNGLADSDKARAYLDRRGFRPMSPELATAAVTQAIADTERPYVVIADIDWSKI 1946
Db 2478 IAWGTWQGNGLA-TGQVSEHLRRRGMFAMPPELAVTAVDGAIA-SGRPSLLVADIDWKKL 2535
Qy 1947 EH--TSQTSDLV-----SAARERERPAVQRPTPPAELHKTLAHQTSADQRAALLELVRDHV 1999
Db 2536 GPVLSSKSSVLLEDLPQAQGTEEARSTVEQTESTNLRQLLMGRSRSEQEEELLSLVRIHS 2595
Qy 2000 AAVLRHADPKAIAPDQSFRALGFDSLTAVEFRNLLIKATGLRLPVSLVFDHPTPAKLAVH 2059
Db 2596 AAVLGRDDSEAIPPGRLFRDLGFDSLAAVELRNHLAAQTELALPTTLVFDYPSPTKLAQF 2655
Qy 2060 LQNQLRGTAAESAP--SAAAVTAEASVTEPIAIVGMACRFPGGVTSADDFWDLISSEQDA 2117
Db 2656 LLSEI----AEFQPDNSTPLPRPRAELDEPIAIVGMACRFPGGVTSADDFWDLISSEQDA 2711
Qy 2118 IGGFPTDRGWDLDTLYDPDPDHPGTCYTRNGGFLYDAGHFDAEFFGISPREALAMDPQQR 2177
Db 2712 IGGFPTDRGWDLDTLYDPDPDHPGTCYTRNGGFLYDAGHFDAEFFGISPREALAMDPQQR 2771
Qy 2178 LLLETAWETIEHAGINPHTLHGTPTGVFTGTNGQDYALRVHNAGQSTDGFALTGTAGSVI 2237
Db 2772 LLLETAWETIEHAGINPHTLHGTPTGVFTGTNGQDHAAHIRQAPSGTEGFVLTGAATSIA 2831
Qy 2238 SGRISYTFGFEGPAVSVDTACSSSLVALHLACQALRAGECSMALAGGVTVMSSPGAFVEF 2297
Db 2832 SGRISYILGLEGPAVTLDTACSSSLVALHLACQSLRSGECTMALAGGATVMTTPITFTEF 2891
Qy 2298 SRQRGLAADGHCKAFSAAADGTGWGEGVGMLLVERLSDAHRNGHRVLAVVRGSAVNQDGA 2357
Db 2892 ARQRGLAPDGRCKAFSAAADGTGWGEGVGMLLVERLSDARRNGHRVLAVVRGSAVNQDGA 2951

```
Qy       2358 SNGLTAPNGPSQQRVIRQALANAGLSAGDVDAVEAHGTGTTLGDPIEAQALLATYGQDRA 2417
Db       2952 SNGLTAPNGPSQQRVIRQALANADLTPADVDAVEAHGTGTTLGDPIEAQAILATYGQDRP 3011

Qy       2418 GEGPLWLGSVKSNVGHTQAAAGVAGVIKMVMALRHGLLPRTLHVDEPSPHVDWSAGAVQL 2477
Db       3012 GNGPLWLGSVKSNVGHTQAAAGVAGVIKMVMALRHRTLPPTLHADEPSPHVDWSAGAVQL 3071

Qy       2478 LTETVPWPGGEGRLRRAGVSSFGVSGTNAHVILEEAPADDVPGGPPAGEGDAGSDDEAAA 2537
Db       3072 LTETVPWPGGEGRPRRAGVSSFGVSGTNAHVILEEAPADDVPGGPPADE-DAGSGEEAAA 3130

Qy       2538 GSPGVWPWLVSAKSQPALRAQAQALHAHLTDHPGLDLADVGYTLAHARAVFDHRATLIAA 2597
Db       3131 GSPGVWPWLVSAKSQPALRAQAQALHAHLTDHPGLDLADVGYTLAHARAVFDHRATLIAA 3190

Qy       2598 DRDTFLQALQALAAGEPHPAVIHSSAPGGTGTGEAAGKTAFICSGQGTQRPGMAHGLYHT 2657
Db       3191 DRDTFLQALQALAAGEPHPAVIHSSAPGGTGTGEAAGKTAFICSGQGTQRPGMAHGLYHT 3250

Qy       2658 HPVFAAALNDICTHLDPHLDHPLLPLLTQ--NDNDNEDAAALLQQTRYAQPALFAFQVAL 2715
Db       3251 HPVFAAALNDICTHLDPHLDHPLLPLLTQNDNDNDNEDAAALLQQTPYAQPALFAFQVAL 3310

Qy       2716 HRLLTDGYHITPHYYAGHSLGEITAAHLAGILTLTDATTLITQRATLMQTMPPGTMTTLH 2775
Db       3311 HRLLTDGYHITPHYYAGHSLGEITAAHLAGILTLTDATTLITQRATLMQTMPPGTMTTLH 3370

Qy       2776 TTPHHITHHLTAHENDLAIAAINTPTSLVISGTPHTVQHITTLCQQQGIKTKTLPTNHAF 2835
Db       3371 TTPHHITHHLTAHENDLAIAAINTPTSLVISGTPHTVQHITTLCQQQGIKTKTLPTNHAF 3430

Qy       2836 HSPHTNPILNQLHQHTQTLTYHPPHTPLITANTPPDQLLTPHYWTQQARNTVDYATTTQT 2895
Db       3431 HSPHTNPILNQLHQHTQTLTYHPPHTPLITANTPPDQLLTPHYWTQQARNTVDYATTTQT 3490

Qy       2896 LHQHGVTTYIELGPDNTLTTLTHHNLPNPPTTTLTLTHPHHHPQTHLLTNLAKTTTTWHP 2955
Db       3491 LHQHGVTTYIELGPDNTLTTLTHHNLPNTPTTTLTLTHPHHHPQTHLLTNLAKTTTTWHP 3550

Qy       2956 HHYTHHDNQPHTHTHLDLPTYPFQHHHYWLE---------------------------- 2986
Db       3551 HHYTHHHNQPHTHTHLDLPTYPFQHHHYWLELPSAQTSPGQRRSRRSAPDTAESEFWDAV 3610

Qy       2987 ------------------------------------------------------------ 2986
Db       3611 NEEDLQSLAETLDIDASALDTVVPALSAWHRHQHDQARINTWTYQETWKPLTLPTTHQPH 3670

Qy       2987 ------------------------------------------------------------ 2986
Db       3671 QTWLIAIPETQTHHPHITNILTNLHHHGITPIPLTVNHTHTNPQHLHHTLHHTRQQAQNH 3730

Qy       2987 ------------------------------------------------------------ 2986
Db       3731 TTGPITGLLSLLALDETPHPHHPHTPTGTLLNLTLPQTHTQTHPPTPLWYATTNATTTHP 3790

Qy       2987 ----------------------------------------------STQP-------- 2990
Db       3791 NDPLTHPTQAQTWGLARTTLLEHPTHTAGIIDLPTTPTPHTLHHLTQTLTQPHHQTQLAI 3850

Qy       2991 ------------------------------------------------------------ 2990
Db       3851 RTTGTHTRRLTPTTLTPTHQPPTPTPHGTTLITGGTGALATHLTHHLTTHQPTQHLLLTS 3910

Qy       2991 ------------------------------------------------------------ 2990
Db       3911 RTGPHTPHAQHLTTQLQQKGIHLTITTCDTSNPDQLQQLLNTIPPQHPLTTVIHTAGVNL 3970

Qy       2991 ----------------------------------------------GAGN-------- 2994
Db       3971 FAPVSETDAESFSSVTAAKATGAAILHELLLDHETLEHFILFSSGAGAWGSGNQCAYSAA 4030

Qy       2995 ------------------------------------------------------------ 2994
Db       4031 NAYLDALATHRQTHGLPGASIAWGPWAGKGMSAGDAAHGYLEKRGILPMEPRMALAAFHR 4090

Qy       2995 ------------------------------------------------------------ 2994
Db       4091 ARAQRPNSNLIIADIDWERFVPAFTARRHSPLIEDIPEVRQAAQELEAAASTAKTTTAQP 4150

Qy       2995 ------------------------------------------------------------ 2994
Db       4151 IATSLRERLARLTSSKQNQVLLGLIRTGICTVLGLRNPEGIEDQRAFRDLGFDSLTSAQF 4210

Qy       2995 ------------------------------------------------------------ 2994
Db       4211 SKELAKETGLPLPPSLVFDYPTPQECAAHLRTQLVDLDDEEDAALSNALPQVAHRRTVED 4270

Qy       2995 ------------------------------------------------------------ 2994
Db       4271 EPIAIIGMACRFPGGVRSADDLWELLASGKDAIGVFPTDRGWDLDTLYDPDPDHPGTCYT 4330

Qy       2995 ------------------------------------------------------------ 2994
Db       4331 RNGGFLYGAGHFDAEFFGISPREALAMDPQQRLLLETAWETIEHAGINPHTLHGTPTGVF 4390

Qy       2995 ------------------------------------------------------------ 2994
Db       4391 AGINAQDHAAHIRQSRDVETIEGYALTGSSGSVASGRVAYTLGLEGPAVSVDTACSSSLV 4450

Qy       2995 ------------------------------------------------------------ 2994
Db       4451 ALHWAAQALRAGECSMALAGGVTVMSSPGTFVEFSRQRGLAADGRCKAYSAAADGTGWAE 4510

Qy       2995 ------------------------------------------------------------ 2994
Db       4511 GVGMLLVERLSDARRNGHRVLAVVRGSAVNQDGASNGLTAPNGPSQQRVIRQALANAGLT 4570

Qy       2995 ------------------------------------------------------------ 2994
Db       4571 PADVDAVEGHGTGTTLGDPIEAQALLAAYGQHRPHHRPLWLGSLKSNIGHAQAAAGVGGV 4630

Qy       2995 ------------------------------------------------------------ 2994
Db       4631 IKMVMALRNGLLPQTLHVDEPTPQVDWSTGAVQLLTQPVPWPADPAGRPRHAGVSSFGVS 4690

Qy       2995 ------------------------------------------------------------ 2994
Db       4691 GTNAHIILEEAPTPQDSDTDDEPPANAPALPHPLPLPVPVSARSEAGLRAQAQALRQYVA 4750

Qy       2995 ------------------------------------------------------------ 2994
Db       4751 ARPDMSPADIGAGLARGRAVLEHRAVILAADREELAQALTALAAGEPHPHITTGHTRGGD 4810

Qy       2995 ------------------------------------------------------------ 2994
Db       4811 RGGVVFVFPGQGGQWAGMGLTLLTSSPVFAEHIDACEKALTPWVPWSLTDILHRDPDDPA 4870

Qy       2995 ------------------------------------------------------------ 2994
Db       4871 WQQADVVQPVLFSIMVSLAALWRSYGIEPDAVLGHSQGEIAAAHICGALSLKDAAKTVAL 4930

Qy       2995 ------------------------------------------------------------ 2994
Db       4931 RSRALAAVRGRGAMASLPLPAQDVQQLISERWEGQLWVAALNGPHSTTVSGDTKAVDEVL 4990

Qy       2995 ------------------------------------------------------------ 2994
Db       4991 AHCTDTGLRAKRIPVDYASHCPHVQPLHDELLHLLGDITPQPSTVPFFSTVEGTWLDTTT 5050

Qy       2995 ------------------------------------------------------------ 2994
Db       5051 LDAAYWYRNLHQPVRFSHAIQTLTDDGHRAFIEISPHPTLVPAIEDTTENTTENITATGS 5110

Qy       2995 ------------------------------------------------------------ 2994
```

```
Db       5111 LRRGDNDTHRFLTALAHTHTTGIGTPTTWHHHYTQTHPHPNPHTHLDLPTYPFQHQHYWL 5170

Qy       2995 --------VSAAGLDPTEHPLLGATLELATDGGALLAGRLSLRSHPWLADHAVGGTVLLS 3046
                      ::  || || |||| ||| || :   || ||||||:|||| || | | |||
Db       5171 QPPTTTTDLTTTGLTPTHHPLLTATLTLADNNTQLLTGRLSLRTHPWLTDHTVAGMVLLP 5230

Qy       3047 GATPLELALHAGTYVGCDRVDELTLHAPLVVPVDGGVSVQVGVAAADGEGRRLVSVYARG 3106
              |   ||||| ||  | | ||:|||||||||:|    |::|| | |||  | | :::::
Db       5231 GTALLELALQAGERVDCPRVEELTLHAPLVIPHTEDVTLQVTVRAADESGHRALAIHSYS 5290

Qy       3107 GSACGGGGASGGVWTCHASGVLVEAAAGGVVVDG-----LAGVWPPRGAVAVDVDGVRDR 3161
              |:|      ::   || ||:|:|   |       |       | | ||| ||  :::  |  |
Db       5291 GTA----SSADREWTRHATGLLTHHADTDHRADTHTDACLGGSWPPPGAQPIELGDVYGR 5346

Qy       3162 LAG-AGCVLGPVFSGLRAVWRDGGDLLAEVCLPEEAWGD--AAGFGLHPALLDGVVQPLS 3218
              :|  :    |||| || | || | |:|||| |||||  |  || ||:|||||  :   :
Db       5347 MAADSDIAYGPVFQGLHAAWRFGDDVLAEVRLPEEALRDAPAAAFGVHPALLDAALHATA 5406

Qy       3219 VLLPGGTGFGEGAG-------FGEGVRVPAVWGGVSLHRAGVTGVRVRVSAVGRGGGREA 3271
              :   | |  |                |:|  | ||||| || : :|||:|   : |   |
Db       5407 LTPQNGDGSTENVAQESMPDRAAHQARLPFSWSGVSLHTAGSSVLRVRLSRSPQHG--NA 5464

Qy       3272 VSVVVGDEAGVPVASVDRLELRPVDMGQLRAVSVSAGRRGSLYAVQWAEV-----GPVPV 3326
              |::   || | || ::: | ||||   :||| :        ||: : |  |       | |
Db       5465 VALTAADEDGRPVVTIESLALRPVSTEELRAAADRTPEHESLFRLDWVSVPVPANAPSPT 5524

Qy       3327 CGQAWA-------------WHEDV-----------GESGGGPVPGVVVL--------RCP 3354
                : ||                || |                 | | |||:|:         |
Db       5525 ADRPWAVIGAGLPHLPGLTEHEHVTAYDEPADLLLALDRGAPPPGVLVVGGVAHTEAREY 5584

Qy       3355 DAGAGGGGGGGGGGGVGEVV--------------GGVLGVVQGWLGLERFAGSRLVVVTR 3400
              | | |  |         :||                  :|  :| ||| || | |||:|:|
Db       5585 SAEAPGERGTEACEARPDVVHVGVVHTAAVHAAAAQMLARLQAWLGDERLADSRLLVLTC 5644

Qy       3401 GAVVAGPEDGPVDVVGASVWGLVRSAQAEHPDRFVLLDLDTDTGTDLDTGAGAGWGVDGG 3460
              |||      |   |: ||:|||||||||:|||||  |||   :  ||:           : |
Db       5645 GAVARASGDDATDLPGAAVWGLVRSAQSEHPDRITLLDF--ERGTE----------AEPG 5692

Qy       3461 RVAAVVACGEPQLAVRGERLLAARLKRLESSGDVPAQRSGDTRARRSDVPAQRSGGVPA- 3519
              ::|  : ||| |||||    |   || |         | |  |       |||  :  ||:
Db       5693 QLATALNCGERQLAVRPGGLFTPRLVR--------APRVADA------VPAVPAVAVPSA 5738

Qy       3520 -RRSVDVSGREVLPWLSGGSVLVTGGTGVLGAAVARHLAGVCGVRDLLLVSRRGPDAPGA 3578
                :|  :|     |:| ||:||:|||||||  |||||     ||| |||  ||||| ||
Db       5739 GHAAVPAAG----PFLPGGTVLITGGTGVLGRLVARHLVEAHGVRHLLLAGRRGPDAEGA 5794

Qy       3579 EGLRAELAALGAEVRIVACDVGERREVVRLLEGVPAGCPLTGVVHAAGVLDDATIASLTP 3638
                |||||  ||| | :|||| :|:::  ||  :|   |||||||:||:|||  | ||:|
Db       5795 PELRAELGGLGATVEVVACDAADRQQLADLLTRIPDDRPLTGVVHSAGILDDGVITSLSP 5854

Qy       3639 ERLGTVFAAKVDAALLLDELTRGMELSAFVLFSSAAGILGSAGQGNYAAANAALDALAYR 3698
              |||| |  || |||||||||||| ||||||:||||: ::|| ||||||||| || ||:|
Db       5855 ERLGAVLRAKADAALLLDELTRGAELSAFVMFSSASAVVGSPGQGNYAAANAVLDFLAHR 5914

Qy       3699 RRAAGLPGVSLAWGLWEEASGMTGHLAGTDHRRIIRSGLHPMSTPDALALFDAALALDRP 3758
              ||| ||| |||||||||| :||||||    || || |:|: |: | :||||||||||    |
Db       5915 RRAEGLPAVSLAWGLWEEGTGMTGHLDVDDHARISRAGMRPLPTAEALALFDAALADGEP 5974

Qy       3759 VLLPADL--------RPAPPLPPLLQDLLPATRRRTTRTTTTGG---ADNGAQLHARLAG 3807
              |:|| |          : |:||||| ||   | |:        |   || |   |||
Db       5975 FLMPARLDLTAVRSGAASAPVPPLLQGLLQLPRSRSAAAAPGHGAPAADEAAAWRERLAR 6034

Qy       3808 QTHEQQHTTLLALVRSHIATVLGHTTPDTIPPDRAFRDLGFDSLTAVELRNRLSRTTGLR 3867
              |:  ::   || |||||:| ||||:  | |   ||||:||||||||||||||:  ||||
Db       6035 QSAGERRQALLRLVRSHVAAVLGHSGADGIDASRAFRELGFDSLTAVELRNRLTAATGLR 6094

Qy       3868 LPTTLAFDHPNPTTLTHHLHTQLQPQPD 3895
              |  ||||| | |   |  ||  :| |  :
Db       6095 LRATLAFDFPTPAALAEHLGERLLPDQE 6122
```

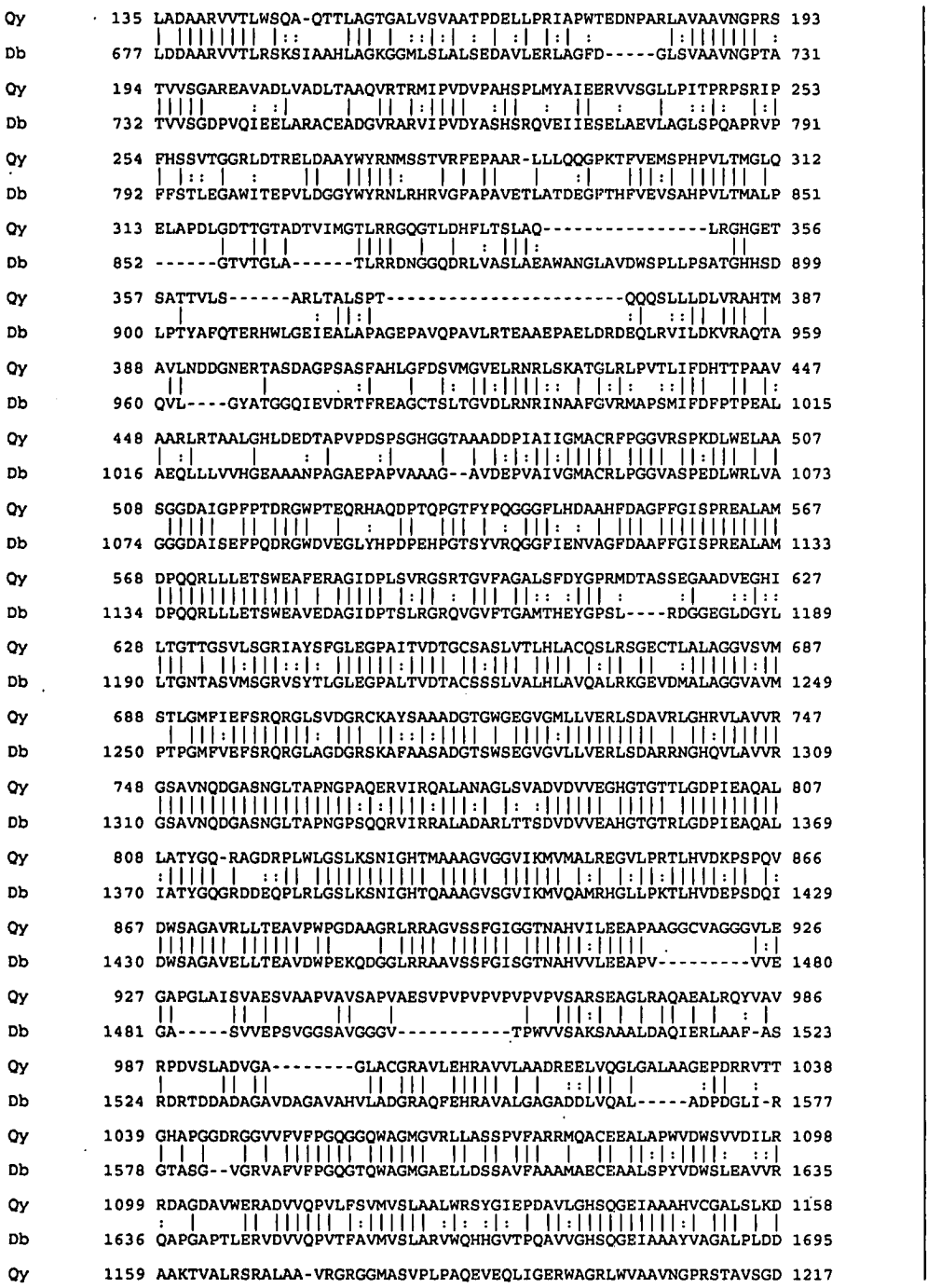
RESULT 8

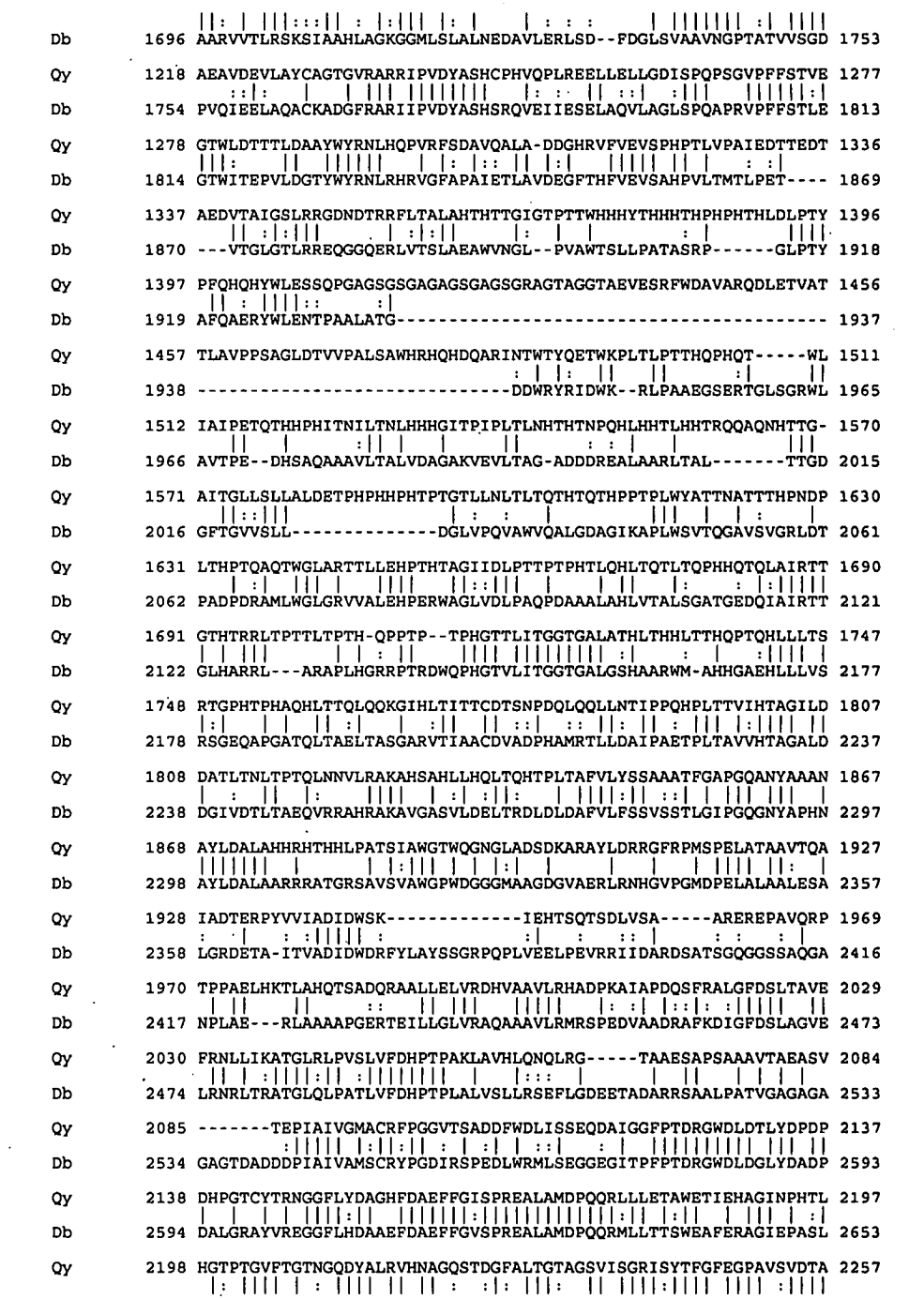
```
AAB18637
ID   AAB18637 standard; Protein; 4551 AA.
XX
AC   AAB18637;
XX
DT   22-JAN-2001  (first entry)
XX
DE   Amino acid sequence of narbonolide synthase subunit 1 (PICAI).
XX
KW   Narbonolide synthase; polyketide synthase gene; narbonolide polyketide;
KW   antibiotic; C12-hydroxyalse; picK; desosamine biosynthesis;
KW   desosaminyl transferase enzyme; ketolide; beta-glucosidase enzyme;
KW   picromycin biosynthesis.
XX
OS   Streptomyces venezuelae.
XX
PN   US6117659-A.
XX
PD   12-SEP-2000.
XX
PF   27-MAY-1999;   99US-0320878.
XX
PR   28-MAY-1998;   98US-0087080.
PR   22-SEP-1998;   98US-0100880.
PR   08-FEB-1999;   99US-0119139.
PR   20-MAY-1999;   99US-0134990.
PR   30-APR-1997;   97US-0846247.
PR   06-MAY-1998;   98US-0073538.
PR   28-AUG-1998;   98US-0141908.
XX
PA   (KOSA-) KOSAN BIOSCIENCES INC.
XX
PI   Ashley G,  Betlach MC,  Betlach M,  Tang L,  McDaniel R;
XX
DR   WPI; 2000-610844/58.
XX
PT   New recombinant picK hydroxylase gene of Streptomyces venezuelae useful
PT   for converting ketolides to antibiotics and as antibiotics and
PT   intermediates in the synthesis of compounds with pharmaceutical value
PT   -
XX
PS   Disclosure; Columns 9-10; 117pp; English.
XX
CC   The present sequence represents a narbonolide synthase subunit 1
CC   (PICAI). The nucleotide sequence encoding it is used in the course of
CC   the invention. The specification describes a recombinant DNA compound
CC   expressing recombinant polyketide synthase genes in host cells for the
CC   production of narbonolide, narbonolide derivatives and polyketides that
CC   are useful as antibiotics and as intermediates in the synthesis of
CC   compounds with pharmaceutical value. The DNA compounds may also encode
CC   a C12-hydroxyalse (picK), desosamine biosynthesis and desosaminyl
CC   transferase enzymes (useful for conversion of ketolides to antibiotics),
CC   and the beta-glucosidase enzyme (involved in picromycin biosynthesis).
CC   These compounds are also useful for increasing the antibiotic activity
CC   of a compound relative to the unhydroxylated compound. The recombinant
CC   host cells are useful as genetic systems that allow rapid engineering
CC   of the narbonolide polyketide synthase. These would be valuable for
CC   creating novel ketolide analogs for pharmaceutical applications.
XX
SQ   Sequence   4551 AA;
```

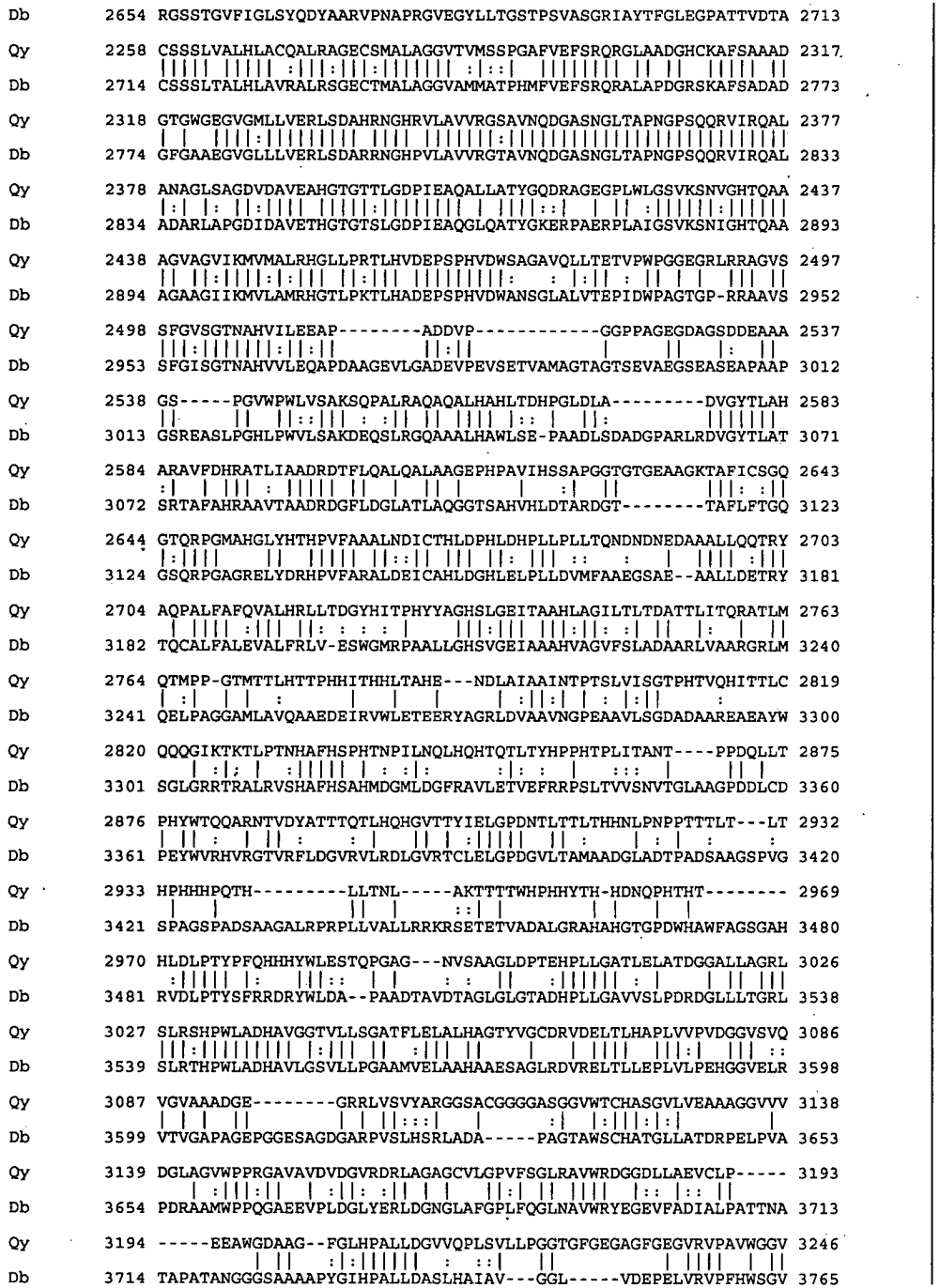
Query Match 39.1%; Score 8151; DB 21; Length 4551;
Best Local Similarity 44.8%; Pred. No. 0;
Matches 1910; Conservative 462; Mismatches 1307; Indels 586; Gaps 94;

```
Qy         15 GEVLGVADEADGGVVFVFPGQGPQWPGMGRELLDASDVFRESVRACEAAFAPYVDWSVEQ 74
                | | |||    | | ||||||| || ||| ||||:| ||  ::  |||| :||||||:|
Db        558 GLVRGVASGV-GRVAFVFPGQGTQWAGMGAELLDSSAVFAAAMAECEAALSPYVDWSLEA 616

Qy         75 VLRDSPDAPGLDRVDVVQPTLFAVMISLAALWRSQGVEPCAVLGHSLGEIAAAHVSGGLS 134
                |:| :| || |:|||||||  ||||:||| :|:  || | ||:||| ||||||:|:| ||
Db        617 VVRQAPGAPTLERVDVVQPVTFAVMVSLARVWQHHGVTPQAVVGHSQGEIAAAYVAGALS 676
```







RESULT 9
AAY67201
ID AAY67201 standard; protein; 4551 AA.
XX
AC AAY67201;
XX
DT 23-MAR-2000 (first entry)
XX
DE Narbonolide synthase subunit 1 (PICAI) protein sequence.
XX
KW Narbonolide polyketide synthase; PKS; narbonolide synthase subunit 1;

```
KW   PICAI; antibiotic production; narbomycin; picromycin; ketolide.
XX
OS   Streptomyces venezuelae.
XX
PN   WO9961599-A2.
XX
PD   02-DEC-1999.
XX
PF   27-MAY-1999;   99WO-US11814.
XX
PR   28-MAY-1998;   98US-0087080.
PR   28-AUG-1998;   98US-0141908.
PR   22-SEP-1998;   98US-0100880.
PR   08-FEB-1999;   99US-0119139.
XX
PA   (KOSA-) KOSAN BIOSCIENCES INC.
XX
PI   Ashley G,  Betlach MC,  Betlach M,  McDaniel R,  Tang L;
XX
DR   WPI; 2000-072618/06.
DR   N-PSDB; AAZ56001.
XX
PT   New recombinant DNA encoding a domain of narbonolide polyketide
PT   synthase, for production of ketolide antibiotics -
XX
PS   Example 2; Page 11-12; 98pp; English.
XX
CC   This is the Streptomyces venezuelae narbonolide synthase subunit 1,
CC   PICAI protein sequence. The invention relates to recombinant DNA
CC   containing a coding sequence for a narbonolide polyketide synthase
CC   (PKS). Polyketides are compounds synthesised from 2-carbon units through
CC   a series of condensations and subsequent modifications. Modular PKSs are
CC   responsible for the production of many antibiotics including picromycin.
CC   The narbonolide PKS consists of a loading module, six extender modules,
CC   and two thioester domains. Four proteins make up the narbonolide PKS
CC   (PICAI, PICAII, PICAIII and PICAIV). PICAI includes the loading module
CC   and extender modules 1 and 2, PICAII includes extender modules 3 and 4,
CC   PICAIII includes extender module 5 and PICAIV includes extender module 6
CC   and a type II thioesterase domain. The second type II thioesterase
CC   domain is found on the PICB protein. The nucleotide sequences encoding
CC   all of these proteins can be isolated in recombinant form from the
CC   recombinant cosmid pKOS023-27 (see AAZ56001). Narbonolide is
CC   desosaminylated in S. venezuelae to yield narbomycin, the desosaminyl
CC   transferase enzyme is required for this  conversion, and the desosamine
CC   biosynthetic genes are also found in cosmid pKOS023-27. The recombinant
CC   DNA of the invention is used to express, in transformed cells,
CC   narbonolide (or its derivatives) or other ketolides (particularly
CC   hybrids), which may then be converted (e.g. by other enzymes
CC   recombinantly expressed in the same hosts) to polyketide antibiotics or
CC   their intermediates. The antibiotics are useful in human or veterinary
CC   medicine.
XX
SQ   Sequence   4551 AA;

  Query Match             39.1%;  Score 8151;  DB 21;  Length 4551;
  Best Local Similarity   44.8%;  Pred. No. 0;
  Matches 1910;  Conservative  462;  Mismatches 1307;  Indels  586;  Gaps   94;

Qy         15 GEVLGVADEADGGVVFVFPGQGPQWPGMGRELLDASDVFRESVRACEAAFAPYVDWSVEQ  74
Db        558 GLVRGVASGV-GRVAFVFPGQGTQWAGMGAELLDSSAVFAAAMAECEAALSPYVDWSLEA  616

Qy         75 VLRDSPDAPGLDRVDVVQPTLFAVMISLAALWRSQGVEPCAVLGHSLGEIAAAHVSGGLS  134
Db        617 VVRQAPGAPTLERVDVVQPVTFAVMVSLARVWQHHGVTPQAVVGHSQGEIAAAYVAGALS  676

Qy        135 LADAARVVTLWSQA-QTTLAGTGALVSVAATPDELLPRIAPWTEDNPARLAVAAVNGPRS  193
Db        677 LDDAARVVTLRSKSIAAHLAGKGGMLSLALSEDAVLERLAGFD-----GLSVAAVNGPTA  731

Qy        194 TVVSGAREAVADLVADLTAAQVRTRMIPVDVPAHSPLMYAIEERVVSGLLPITPRPSRIP  253
Db        732 TVVSGDPVQIEELARACEADGVRARVIPVDYASHSRQVEIIESELAEVLAGLSPQAPRVP  791

Qy        254 FHSSVTGGRLDTRELDAAYWYRNMSSTVRFEPAAR-LLLQQGPKTFVEMSPHPVLTMGLQ  312
Db        792 FFSTLEGAWITEPVLDGGYWYRNLRHRVGFAPAVETLATDEGFTHFVEVSAHPVLTMALP  851

Qy        313 ELAPDLGDTTGTADTVIMGTLRRGQGTLDHFLTSLAQ---------------LRGHGET  356
Db        852 ------GTVTGLA------TLRRDNGGQDRLVASLAEAWANGLAVDWSPLLPSATGHHSD  899

Qy        357 SATTVLS------ARLTALSPT----------------------QQQSLLLDLVRAHTM  387
Db        900 LPTYAFQTERHWLGEIEALAPAGEPAVQPAVLRTEAAEPAELDRDEQLRVILDKVRAQTA  959

Qy        388 AVLNDDGNERTASDAGPSASFAHLGFDSVMGVELRNRLSKATGLRLPVTLIFDHTTPAAV  447
Db        960 QVL----GYATGGQIEVDRTFREAGCTSLTGVDLRNRINAAFGVRMAPSMIFDFPTPEAL 1015

Qy        448 AARLRTAALGHLDEDTAPVPDSPSGHGGTAAADDPIAIIGMACRFPGGVRSPKDLWELAA  507
Db       1016 AEQLLLVVHGEAAANPAGAEPAPVAAAG--AVDEPVAIVGMACRLPGGVASPEDLWRLVA 1073

Qy        508 SGGDAIGPFPTDRGWPTEQRHAQDPTQPGTFYPQGGGFLHDAAHFDAGFFGISPREALAM  567
Db       1074 GGGDAISEFPQDRGWDVEGLYHPDPEHPGTSYVRQGGFIENVAGFDAAFFGISPREALAM 1133

Qy        568 DPQQRLLLETSWEAFERAGIDPLSVRGSRTGVFAGALSFDYGPRMDTASSEGAADVEGHI  627
Db       1134 DPQQRLLLETSWEAVEDAGIDPTSLRGRQVGVFTGAMTHEYGPSL----RDGGEGLDGYL 1189

Qy        628 LTGTTGSVLSGRIAYSFGLEGPAITVDTGCSASLVTLHLACQSLRSGECTLALAGGVSVM  687
Db       1190 LTGNTASVMSGRVSYTLGLEGPALTVDTACSSSLVALHLAVQALRKGEVDMALAGGVAVM 1249

Qy        688 STLGMFIEFSRQRGLSVDGRCKAYSAAADGTGWGEGVGMLLVERLSDAVRLGHRVLAVVR  747
Db       1250 PTPGMFVEFSRQRGLAGDGRSKAFAASADGTSWSEGVGVLLVERLSDARRNGHQVLAVVR 1309

Qy        748 GSAVNQDGASNGLTAPNGPAQERVIRQALANAGLSVADVDVVEGHGTGTTLGDPIEAQAL  807
Db       1310 GSAVNQDGASNGLTAPNGPSQQRVIRRALADARLTTSDVDVVEAHGTGTRLGDPIEAQAL 1369

Qy        808 LATYGQ-RAGDRPLWLGSLKSNIGHTMAAAGVGGVIKMVMALREGVLPRTLHVDKPSPQV  866
Db       1370 IATYGQGRDDEQPLRLGSLKSNIGHTQAAAGVSGVIKMVQAMRHGLLPKTLHVDEPSDQI 1429

Qy        867 DWSAGAVRLLTEAVPWPGDAAGRLRRAGVSSFGIGGTNAHVILEEAPAAGGCVAGGGVLE  926
Db       1430 DWSAGAVELLTEAVDWPEKQDGGLRRAAVSSFGISGTNAHVVLEEAPV---------VVE 1480

Qy        927 GAPGLAISVAESVAAPVAVSAPVAESVPVPVPVPVPVPVSARSEAGLRAQAEALRQYVAV  986
Db       1481 GA-----SVVEPSVGGSAVGGGV-----------TPWVVSAKSAAALDAQIERLAAF-AS 1523

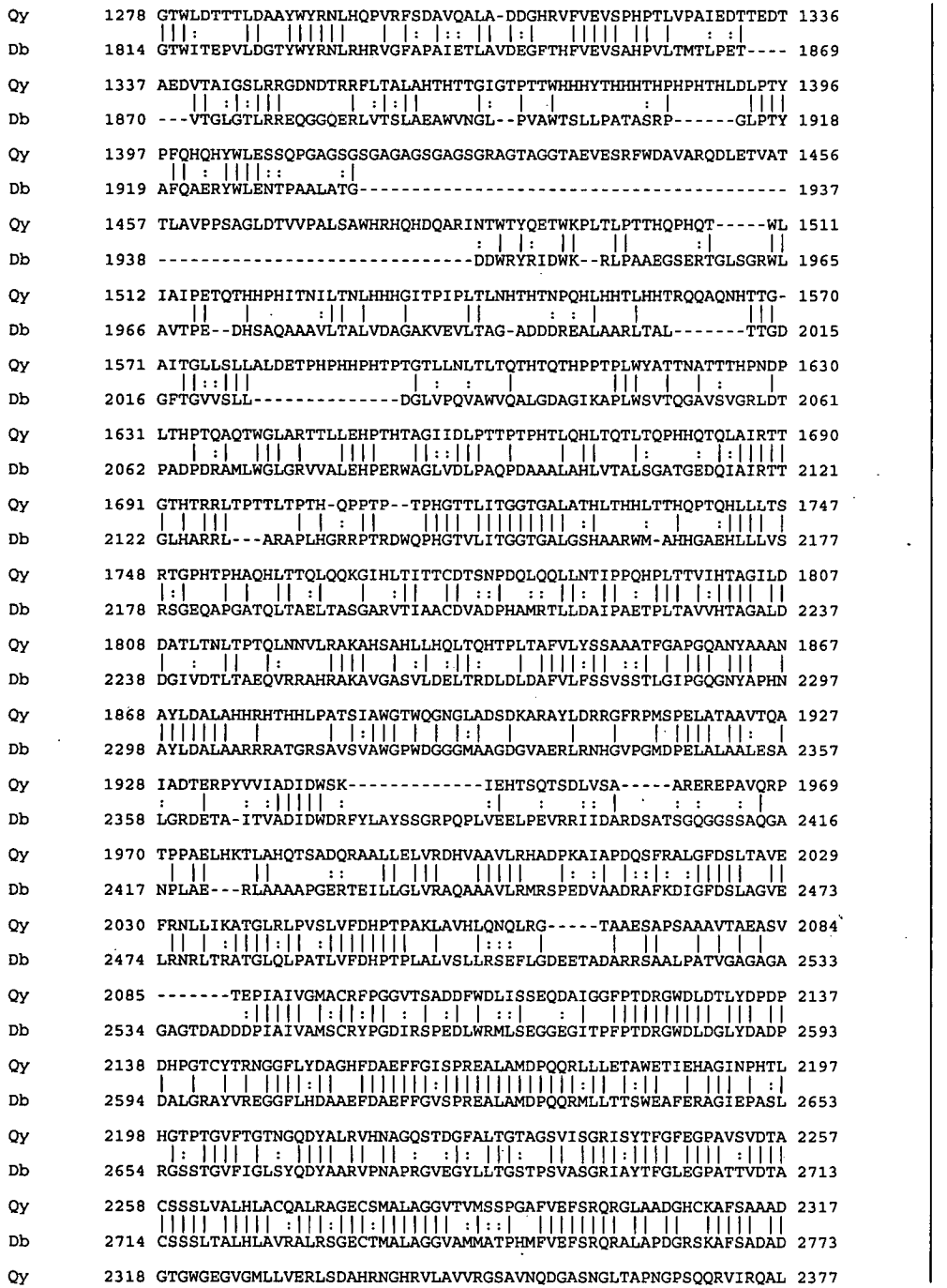
Qy        987 RPDVSLADVGA--------GLACGRAVLEHRAVVLAADREELVQGLGALAAGEPDRRVTT 1038
Db       1524 RDRTDDADAGAVDAGAVAHVLADGRAQFEHRAVALGAGADDLVQAL-----ADPDGLI-R 1577

Qy       1039 GHAPGGDRGGVVFVFPGQGGQWAGMGVRLLASSPVFARRMQACEEALAPWVDWSVVDILR 1098
Db       1578 GTASG--VGRVAFVFPGQGTQWAGMGAELLDSSAVFAAAMAECEAALSPYVDWSLEAVVR 1635

Qy       1099 RDAGDAVWERADVVQPVLFSVMVSLAALWRSYGIEPDAVLGHSQGEIAAAHVCGALSLKD 1158
Db       1636 QAPGAPTLERVDVVQPVTFAVMVSLARVWQHHGVTPQAVVGHSQGEIAAAYVAGALPLDD 1695

Qy       1159 AAKTVALRSRALAA-VRGRGGMASVPLPAQEVEQLIGERWAGRLWVAAVNGPRSTAVSGD 1217
Db       1696 AARVVTLRSKSIAAHLAGKGGMLSLALNEDAVLERLSD--FDGLSVAAVNGPTATVVSGD 1753

Qy       1218 AEAVDEVLAYCAGTGVRARRIPVDYASHCPHVQPLREELLELLGDISPQPSGVPFFSTVE 1277
Db       1754 PVQIEELAQACKADGFRARIIPVDYASHSRQVEIIESELAQVLAGLSPQAPRVPFFSTLE 1813
```



Qy 1278 GTWLDTTTLDAAYWYRNLHQPVRFSDAVQALA-DDGHRVFVEVSPHPTLVPAIEDTTEDT 1336
Db 1814 GTWITEPVLDGTYWYRNLRHRVGFAPAIETLAVDEGFTHFVEVSAHPVLTMTLPET---- 1869

Qy 1337 AEDVTAIGSLRRGDNDTRRFLTALAHTHTTGIGTPTTWHHHYTHHHTHPHPHTHLDLPTY 1396
Db 1870 ---VTGLGTLRREQGGQERLVTSLAEAWVNGL--PVAWTSLLPATASRP------GLPTY 1918

Qy 1397 PFQHQHYWLESSQPGAGSGSGAGAGSGAGSGRAGTAGGTAEVESRFWDAVARQDLETVAT 1456
Db 1919 AFQAERYWLENTPAALATG-- 1937

Qy 1457 TLAVPPSAGLDTVVPALSAWHRHQHDQARINTWTYQETWKPLTLPTTHQPHQT-----WL 1511
Db 1938 ----------------------------DDWRYRIDWK--RLPAAEGSERTGLSGRWL 1965

Qy 1512 IAIPETQTHHPHITNILTNLHHHGITPIPLTLNHTHTNPQHLHHTLHHTRQQAQNHTTG- 1570
Db 1966 AVTPE--DHSAQAAAVLTALVDAGAKVEVLTAG-ADDDREALAARLTAL-------TTGD 2015

Qy 1571 AITGLLSLLALDETPHPHHPHTPTGTLLNLTLTQTHTQTHPPTPLWYATTNATTTHPNDP 1630
Db 2016 GFTGVVSLL-------------DGLVPQVAWVQALGDAGIKAPLWSVTQGAVSVGRLDT 2061

Qy 1631 LTHPTQAQTWGLARTTLLEHPTHTAGIIDLPTTPTPHTLQHLTQTLTQPHHQTQLAIRTT 1690
Db 2062 PADPDRAMLWGLGRVVALEHPERWAGLVDLPAQPDAAALAHLVTALSGATGEDQIAIRTT 2121

Qy 1691 GTHTRRLTPTTLTPTH-QPPTP--TPHGTTLITGGTGALATHLTHHLTTHQPTQHLLLTS 1747
Db 2122 GLHARRL---ARAPLHGRRPTRDWQPHGTVLITGGTGALGSHAARWM-AHHGAEHLLLVS 2177

Qy 1748 RTGPHTPHAQHLTTQLQQKGIHLTITTCDTSNPDQLQQLLNTIPPQHPLTTVIHTAGILD 1807
Db 2178 RSGEQAPGATQLTAELTASGARVTIAACDVADPHAMRTLLDAIPAETPLTAVVHTAGALD 2237

Qy 1808 DATLTNLTPTQLNNVLRAKAHSAHLLHQLTQHTPLTAFVLYSSAAATFGAPGQANYAAAN 1867
Db 2238 DGIVDTLTAEQVRRAHRAKAVGASVLDELTRDLDLDAFVLFSSVSSTLGIPGQGNYAPHN 2297

Qy 1868 AYLDALAHHRHTHHLPATSIAWGTWQGNGLADSDKARAYLDRRGFRPMSPELATAAVTQA 1927
Db 2298 AYLDALAARRRATGRSAVSVAWGPWDGGGMAAGDGVAERLRNHGVPGMDPELALAALESA 2357

Qy 1928 IADTERPYVVIADIDWSK-------------IEHTSQTSDLVSA-----ARE-REPAVQRP 1969
Db 2358 LGRDETA-ITVADIDWDRFYLAYSSGRPQPLVEELPEVRRIIDARDSATSGQGGSSAQGA 2416

Qy 1970 TPPAELHKTLAHQTSADQRAALLELVRDHVAAVLRHADPKAIAPDQSFRALGFDSLTAVE 2029
Db 2417 NPLAE---RLAAAAPGERTEILLGLVRAQAAAVLRMRSPEDVAADRAFKDIGFDSLAGVE 2473

Qy 2030 FRNLLIKATGLRLPVSLVFDHPTPAKLAVHLQNQLRG-----TAAESAPSAAAVTAEASV 2084
Db 2474 LRNRLTRATGLQLPATLVFDHPTPLALVSLLRSEFLGDEETADARRSAALPATVGAGAGA 2533

Qy 2085 -------TEPIAIVGMACRFPGGVTSADDFWDLISSEQDAIGGFPTDRGWDLDTLYDPDP 2137
Db 2534 GAGTDADDDPIAIVAMSCRYPGDIRSPEDLWRMLSEGGEGITPFPTDRGWDLDGLYDADP 2593

Qy 2138 DHPGTCYTRNGGFLYDAGHFDAEFFGISPREALAMDPQQRLLLETAWETIEHAGINPHTL 2197
Db 2594 DALGRAYVREGGFLHDAAEFDAEFFGVSPREALAMDPQQRMLLTTSWEAFERAGIEPASL 2653

Qy 2198 HGTPTGVFTGTNGQDYALRVHNAGQSTDGFALTGTAGSVISGRISYTFGFEGPAVSVDTA 2257
Db 2654 RGSSTGVFIGLSYQDYAARVPNAPRGVEGYLLTGSTPSVASGRIAYTFGLEGPATTVDTA 2713

Qy 2258 CSSSLVALHLACQALRAGECSMALAGGVTVMSSPGAFVEFSRQRGLAADGHCKAFSAAAD 2317
Db 2714 CSSSLTALHLAVRALRSGECTMALAGGVAMMATPHMFVEFSRQRALAPDGRSKAFSADAD 2773

Qy 2318 GTGWGEGVGMLLVERLSDAHRNGHRVLAVVRGSAVNQDGASNGLTAPNGPSQQRVIRQAL 2377

Db 2774 GFGAAEGVGLLLVERLSDARRNGHPVLAVVRGTAVNQDGASNGLTAPNGPSQQRVIRQAL 2833

Qy 2378 ANAGLSAGDVDAVEAHGTGTTLGDPIEAQALLATYGQDRAGEGPLWLGSVKSNVGHTQAA 2437
Db 2834 ADARLAPGDIDAVETHGTGTSLGDPIEAQGLQATYGKERPAERPLAIGSVKSNIGHTQAA 2893

Qy 2438 AGVAGVIKMVMALRHGLLPRTLHVDEPSPHVDWSAGAVQLLTETVPWPGGEGRLRRAGVS 2497
Db 2894 AGAAGIIKMVLAMRHGTLPKTLHADEPSPHVDWANSGLALVTEPIDWPAGTGP-RRAAVS 2952

Qy 2498 SFGVSGTNAHVILEEAP--------ADDVP-----------GGPPAGEGDAGSDDEAAA 2537
Db 2953 SFGISGTNAHVVLEQAPDAAGEVLGADEVPEVSETVAMAGTAGTSEVAEGSEASEAPAAP 3012

Qy 2538 GS-----PGVWPWLVSAKSQPALRAQAQALHAHLTDHPGLDLA---------DVGYTLAH 2583
Db 3013 GSREASLPGHLPWVLSAKDEQSLRGQAAALHAWLSE-PAADLSDADGPARLRDVGYTLAT 3071

Qy 2584 ARAVFDHRATLIAADRDTFLQALQALAAGEPHPAVIHSSAPGGTGTGEAAGKTAFICSGQ 2643
Db 3072 SRTAFAHRAAVTAADRDGFLDGLATLAQGGTSAHVHLDTARDGT--------TAFLFTGQ 3123

Qy 2644 GTQRPGMAHGLYHTHPVFAAALNDICTHLDPHLDHPLLPLLTQNDNDNEDAAALLQQTRY 2703
Db 3124 GSQRPGAGRELYDRHPVFARALDEICAHLDGHLELPLLDVMFAAEGSAE--AALLDETRY 3181

Qy 2704 AQPALFAFQVALHRLLTDGYHITPHYYAGHSLGEITAAHLAGILTLTDATTLITQRATLM 2763
Db 3182 TQCALFALEVALFRLV-ESWGMRPAALLGHSVGEIAAAHVAGVFSLADAARLVAARGRLM 3240

Qy 2764 QTMPP-GTMTTLHTTPHHITHHLTAHE---NDLAIAAINTPTSLVISGTPHTVQHITTLC 2819
Db 3241 QELPAGGAMLAVQAAEDEIRVWLETEERYAGRLDVAAVNGPEAAVLSGDADAAREAEAYW 3300

Qy 2820 QQQGIKTKTLPTNHAFHSPHTNPILNQLHQHTQTLTYHPPHTPLITANT----PPDQLLT 2875
Db 3301 SGLGRRTRALRVSHAFHSAHMDGMLDGFRAVLETVEFRRPSLTVVSNVTGLAAGPDDLCD 3360

Qy 2876 PHYWTQQARNTVDYATTTQTLHQHGVTTYIELGPDNTLTTLTHHNLPNPPTTTLT---LT 2932
Db 3361 PEYWVRHVRGTVRFLDGVRVLRDLGVRTCLELGPDGVLTAMAADGLADTPADSAAGSPVG 3420

Qy 2933 HPHHHPQTH---------LLTNL-----AKTTTTWHPHHYTH-HDNQPHTHT-------- 2969
Db 3421 SPAGSPADSAAGALRPRPLLVALLRRKRSETETVADALGRAHAHGTGPDWHAWFAGSGAH 3480

Qy 2970 HLDLPTYPFQHHHYWLESTQPGAG---NVSAAGLDPTEHPLLGATLELATDGGALLAGRL 3026
Db 3481 RVDLPTYSFRRDRYWLDA--PAADTAVDTAGLGLGTADHPLLGAVVSLPDRDGLLLTGRL 3538

Qy 3027 SLRSHPWLADHAVGGTVLLSGATFLELALHAGTYVGCDRVDELTLHAPLVVPVDGGVSVQ 3086
Db 3539 SLRTHPWLADHAVLGSVLLPGAAMVELAAHAAESAGLRDVRELTLLEPLVLPEHGGVELR 3598

Qy 3087 VGVAAADGE--------GRRLVSVYARGGSACGGGGASGGVWTCHASGVLVEAAAGGVVV 3138
Db 3599 VTVGAPAGEPGGGESAGDGARPVSLHSRLADA-----PAGTAWSCHATGLLATDRPELPVA 3653

Qy 3139 DGLAGVWPPRGAVAVDVDGVRDRLAGAGCVLGPVFSGLRAVWRDGGDLLAEVCLP----- 3193
Db 3654 PDRAAMWPPQGAEEVPLDGLYERLDGNGLAFGPLFQGLNAVWRYEGEVFADIALPATTNA 3713

Qy 3194 -----EEAWGDAAG--FGLHPALLDGVVQPLSVLLPGGTGFGEGAGFGEGVRVPAVWGGV 3246
Db 3714 TAPATANGGGSAAAAPYGIHPALLDASLHAIAV---GGL-----VDEPELVRVPFHWSGV 3765

Qy 3247 SLHRAGVTGVRVRVSAVGRGGGREAVSVVVGDEAGVPVASVDRLELRPVDMGQLRAVSVS 3306
Db 3766 TVHAAGAAAARVRLAS----AGTDAVSLSLTDGEGRPLVSVERLTLRPVTADQAAASRVG 3821

Qy 3307 AGRRGSLYAVQW--------------AEVGPVPVCGQ---------------------A 3330

```
Db       3822 ----GLMHRVAWRPYALASSGEQDPHATSYGPTAVLGKDELKVAAALESAGVEVGLYPDL 3877

Qy       3331 WAWHEDVGESGGGPVPGVVVLRCPDAGAGGGGGGGGGGGVGEVVGGVLGVVQGWLGLERF 3390

Db       3878 AALSQDV--AAGAPAPRTVLAPLP-----AGPADGGAEGVRGTVARTLELLQAWLADEHL 3930

Qy       3391 AGSRLVVVTRGAV----VAGPEDGPVDVVGASVWGLVRSAQAEHPDRFVLLDLDTDTGT- 3445

Db       3931 AGTRLLLVTRGAVRDPEGSGADDGGEDLSHAAAWGLVRTAQTENPGRFGLLDLADDASSY 3990

Qy       3446 ----DLDTGAGAGWGVDGGRVAAVVACGEPQLAVRGERLLAARLKRLESSGDVPAQRSGD 3501

Db       3991 RTLPSVLSDAGLR--------------DEPQLALHDGTIRLARL---------------- 4020

Qy       3502 TRARRSDVPAQRSGGVPARRSVDVSGREVLPWLSGGSVLVTGGTGVLGAAVARHLAGVCG 3561

Db       4021 -----ASVRPETGTAAPA----------LAP---EGTVLLTGGTGGLGGLVARHVVGEWG 4062

Qy       3562 VRDLLLVSRRGPDAPGAEGLRAELAALGAEVRIVACDVGERREVVRLLEGVPAGCPLTGV 3621

Db       4063 VRRLLLVSRRGTDAPGADELVHELEALGADVSVAACDVADREALTAVLDAIPAEHPLTAV 4122

Qy       3622 VHAAGVLDDATIASLTPERLGTVFAAKVDAALLLDELTR--GMELSAFVLFSSAAGILGS 3679

Db       4123 VHTAGVLSDGTLPSMTTEDVEHVLRPKVDAAFLLDELTSTPAYDLAAFVMFSSAAAVFGG 4182

Qy       3680 AGQGNYAAANAALDALAYRRRAAGLPGVSLAWGLWEEASGMTGHLAGTDHRRIIRSGLHP 3739

Db       4183 AGQGAYAAANATLDALAWRRRAAGLPALSLGWGLWAETSGMTGELGQADLRRMSRAGIGG 4242

Qy       3740 MSTPDALALFDAALALDR-PVLLP-----ADLRPA-----PPLPPLLQDLLPATRRR--- 3785

Db       4243 ISDAEGIALLDAALRDDRHPVLLPLRLDAAGLRDAAGNDPAGIPALFRDVVGARTVRARP 4302

Qy       3786 ------TTRTT--TTGGADNGAQLHA-----RLAGQTHEQQHTTLLALVRSHIATVLGHT 3832

Db       4303 SAASASTTAGTAGTPGTADGAAETAAVTLADRAATVDGPARQRLLLEFVVGEVAEVLGHA 4362

Qy       3833 TPDTIPPDRAFRDLGFDSLTAVELRNRLSRTTGLRLPTTLAFDHPNPTTLTHHLHTQL-- 3890

Db       4363 RGHRIDAERGFLDLGFDSLTAVELRNRLNSAGGLALPATLVFDHPSPAALASHLDAELPR 4422

Qy       3891 -QPQPDNA-----------------VAPVLAELDKLESALSALDKTDS-ASERVTLRLK 3930

Db       4423 GASDQDGAGNRNGNENGTTASRSTAETDALLAQLTRLEGALVLTGLSDAPGSEEVLEHLR 4482

Qy       3931 SLMLR------------------------WNAPQHPTAESADDD--EKFTSATEAEIFK 3963

Db       4483 SLRSMVTGETGTGTASGAPDGAGSGAEDRPWAAGDGAGGGSEDGAGVPDFMNASAEELFG 4542

Qy       3964 FIDND 3968

Db       4543 LLDQD 4547


RESULT 10
AAY77192
ID   AAY77192 standard; Protein; 4613 AA.
XX
AC   AAY77192;
XX
DT   05-JUN-2000  (first entry)
XX
DE   S. venezuelae macrolide biosynthetic enzyme PikAI, SEQ ID NO:31.
XX
KW   Desosamine biosynthesis; macrolide; polyketide; methymycin; pikromycin;
KW   neomethymycin; narbomycin; polyhydroxyalkanoate monomer synthase;
KW   biopolymer; antibiotic; chemotherapeutic; immunosuppressant; asthma;
KW   chronic obstructive pulmonary disease; respiratory inflammation;
KW   hypercholesterolaemia; crop protection agent.
XX
OS   Streptomyces venezuelae ATCC15439.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 1386
FT                   /note= "Encoded by GNN"
FT   Misc-difference 4581
FT                   /note= "Encoded by GNN"
XX
PN   WO200000620-A2.
XX
PD   06-JAN-2000.
XX
PF   25-JUN-1999;   99WO-US14398.
XX
PR   26-JUN-1998;   98US-0105537.
XX
PA   (MINU ) UNIV MINNESOTA.
XX
PI   Sherman DH,  Liu H,  Xue Y,  Zhao L;
XX
DR   WPI; 2000-160679/14.
DR   N-PSDB; AAZ87297.
XX
PT   Desosamine and macrolide biosynthetic gene clusters, useful for, e .g.
PT   synthesis of methymycin and pikromycin -
XX
PS   Claim 19; Page 383-398; 438pp; English.
XX
CC   The invention relates to an isolated and purified nucleic acid segment
CC   comprising a desosamine biosynthetic gene cluster, a fragment or its
CC   biologically active variant, where the nucleic acid sequence is not
CC   derived from the eryC gene cluster of Saccharopolyspora erythraea or
CC   Streptomyces antibioticus. The invention also relates to a macrolide
CC   biosynthetic gene cluster, or fragments thereof. The macrolide
CC   biosynthetic gene cluster encodes proteins which synthesise methymycin,
CC   pikromycin, neomethymycin, narbomycin or a combination of these
CC   compounds. Recombinant or augmented cells comprising the desosamine
CC   and/or macrolide biosynthetic gene clusters are useful for the
CC   production of biologically active macrolides. The macrolide biosynthetic
CC   proteins are useful for synthesis of methymycin, pikromycin,
CC   neomethymycin and narbomycin. The alternative termination of polyketide
CC   synthesis may be useful to prepare novel antibiotics and
CC   polyhydroxyalkanoate (PHA) monomers. The compounds produced by the
CC   recombinant host cells are useful as biopolymers, e.g., in packaging or
CC   biomedical applications, to engineer PHA monomer synthases or to prepare
CC   biologically active agents, such as chemotherapeutics,
CC   immunosuppressants, agents to treat asthma, chronic obstructive pulmonary
CC   disease as well as other diseases involving respiratory inflammation,
CC   cholesterol-lowering agents or macrolide-based antibiotics which are
CC   active against a variety of organisms, e.g., bacteria, including
CC   multi-drug resistant pneumococci and other respiratory pathogens, as well
CC   as viral parasitic pathogens, or as crop protection agents (e.g.,
CC   fungicides or insecticides) via expression of polyketides in plants.
CC   Sequences AAY77190-Y77197 represent macrolide biosynthetic enzymes from
CC   Streptomyces venezuelae ATCC 15439, which are encoded by sequences
CC   AAZ87295-Z87302.
XX
SQ   Sequence   4613 AA;

  Query Match           39.1%;  Score 8148;  DB 21;  Length 4613;
  Best Local Similarity   44.8%;  Pred. No. 0;
  Matches 1909;  Conservative  463;  Mismatches 1307;  Indels  586;  Gaps   94;

Qy         15 GEVLGVADEADGGVVFVFPGQGPQWPGMGRELLDASDVFRESVRACEAAFAPYVDWSVEQ 74

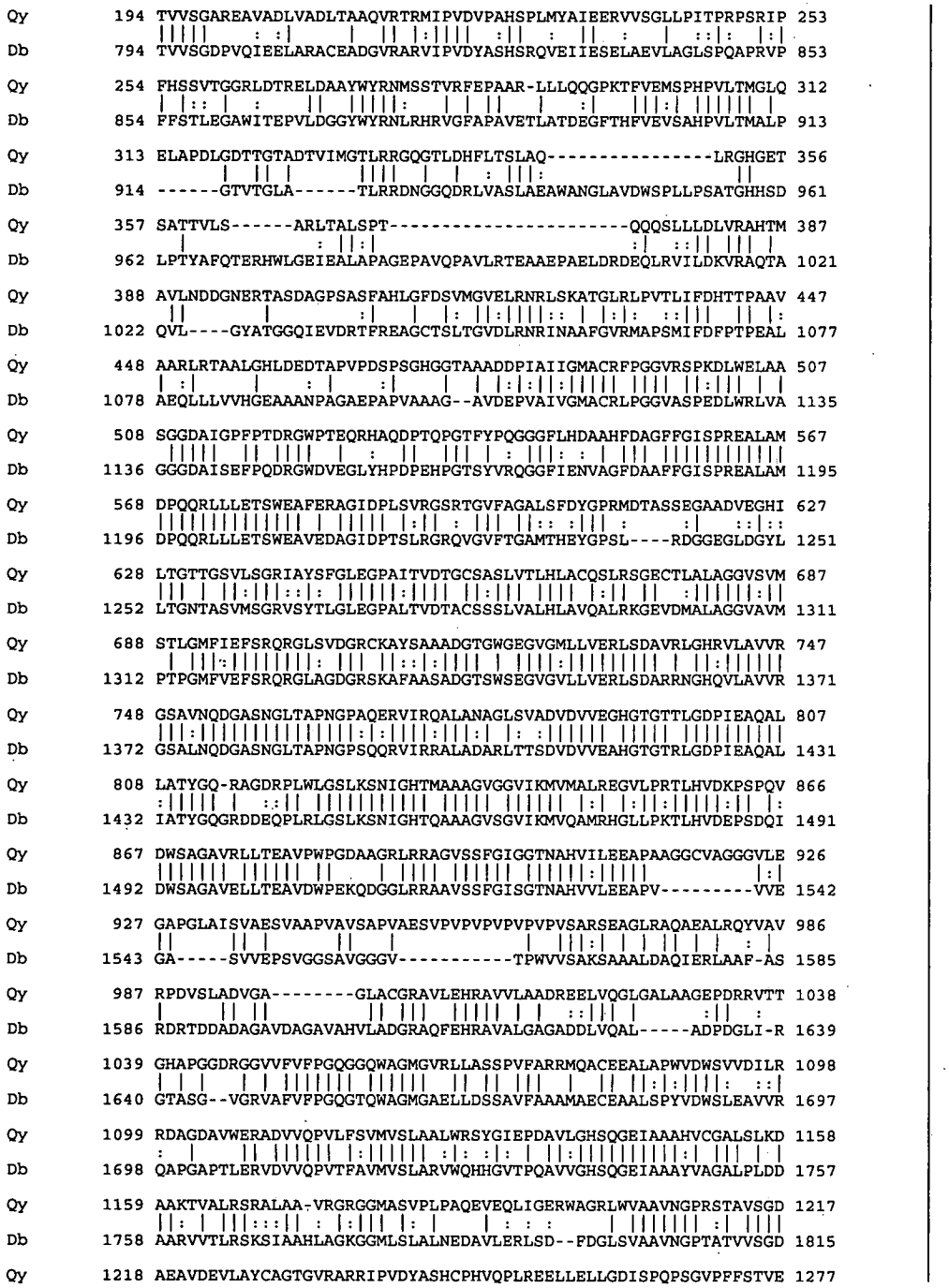
Db        620 GLVRGVASGV-GRVAFVFPGQGTQWAGMGAELLDSSAVFAAAMAECEAALSPYVDWSLEA 678

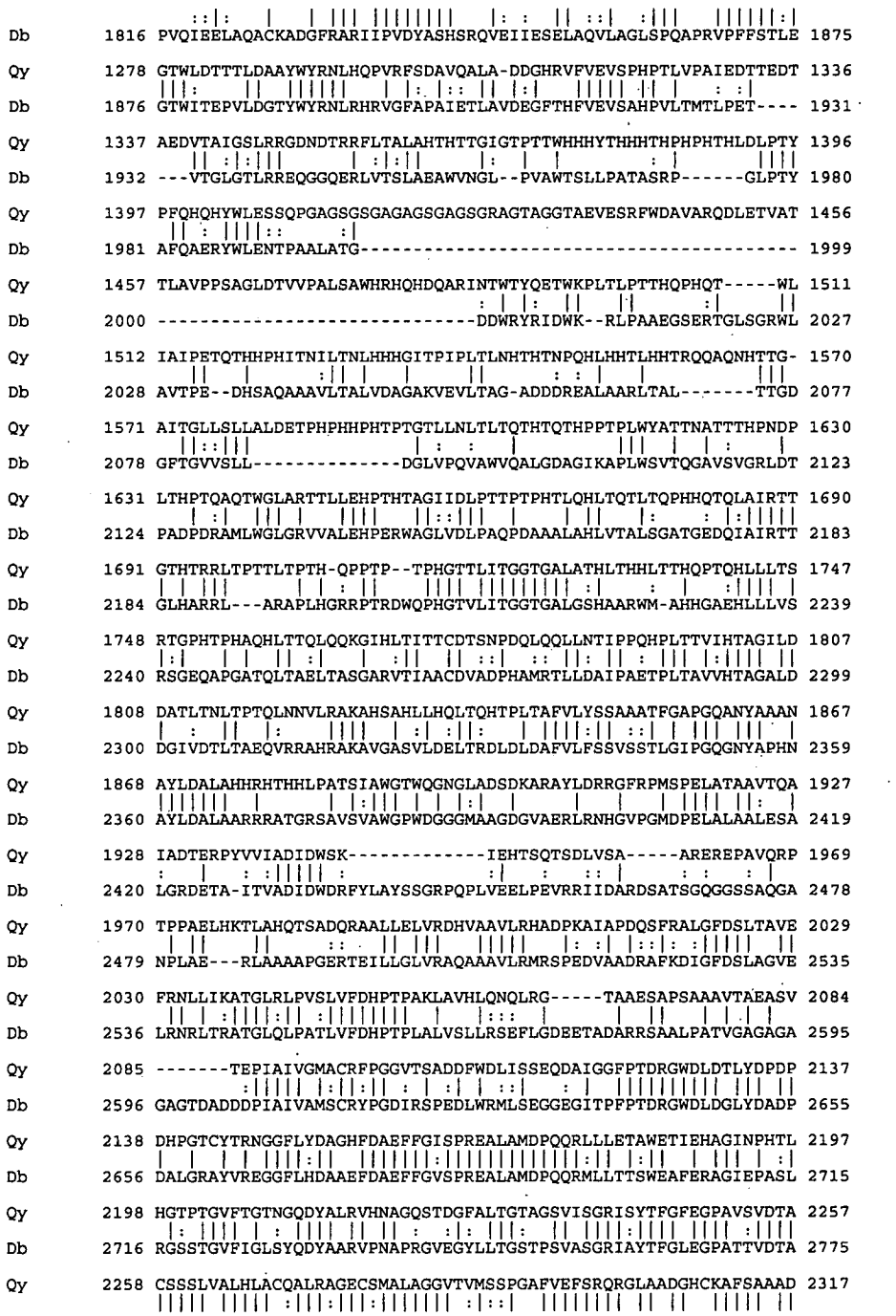
Qy         75 VLRDSPDAPGLDRVDVVQPTLFAVMISLAALWRSQGVEPCAVLGHSLGEIAAAHVSGGLS 134

Db        679 VVRQAPGAPTLERVDVVQPVTFAVMVSLARVWQHHGVTPQAVVGHSQGEIAAAYVAGALS 738

Qy        135 LADAARVVTLWSQA-QTTLAGTGALVSVAATPDELLPRIAPWTEDNPARLAVAAVNGPRS 193

Db        739 LDDAARVVTLRSKSIAAHLAGKGGMLSLALSEDAVLERLAGFD-----GLSVAAVNGPTA 793
```





```
Db       2776 CSSSLTALHLAVRALRSGECTMALAGGVAMMATPHMFVEFSRQRALAPDGRSKAFSADAD 2835

Qy       2318 GTGWGEGVGMLLVERLSDAHRNGHRVLAVVRGSAVNQDGASNGLTAPNGPSQQRVIRQAL 2377
             | |  ||||:|||||||| |||| |||||||:|||||||||||||||||||||||||||||
Db       2836 GFGAAEGVGLLLVERLSDARRNGHPVLAVVRGTAVNQDGASNGLTAPNGPSQQRVIRQAL 2895

Qy       2378 ANAGLSAGDVDAVEAHGTGTTLGDPIEAQALLATYGQDRAGEGPLWLGSVKSNVGHTQAA 2437
             |:| |: ||:|||| |||||:|||||||| | ||||::|  | || :||||||:||||||
Db       2896 ADARLAPGDIDAVETHGTGTSLGDPIEAQGLQATYGKERPAERPLAIGSVKSNIGHTQAA 2955

Qy       2438 AGVAGVIKMVMALRHGLLPRTLHVDEPSPHVDWSAGAVQLLTETVPWPGGEGRLRRAGVS 2497
             || ||:||||:|:||| ||:||| |||||||||:   : |:|| : || | |  ||| ||
Db       2956 AGAAGIIKMVLAMRHGTLPKTLHADEPSPHVDWANSGLALVTEPIDWPAGTGP-RRAAVS 3014

Qy       2498 SFGVSGTNAHVILEEAP--------ADDVP------------GGPPAGEGDAGSDDEAAA 2537
             |||:|||||||:||:||        ||:||            |      ||   |:  ||
Db       3015 SFGISGTNAHVVLEQAPDAAGEVLGADEVPEVSETVAMAGTAGTSEVAEGSEASEAPAAP 3074

Qy       2538 GS-----PGVWPWLVSAKSQPALRAQAQALHAHLTDHPGLDLA---------DVGYTLAH 2583
             ||     ||  ||::||| : :|| || |||| |:: |  ||:          |||||||
Db       3075 GSREASLPGHLPWVLSAKDEQSLRGQAAALHAWLSE-PAADLSDADGPARLRDVGYTLAT 3133

Qy       2584 ARAVFDHRATLIAADRDTFLQALQALAAGEPHPAVIHSSAPGGTGTGEAAGKTAFICSGQ 2643
             :|  | ||| : ||||| ||  |  || |      |  :|  ||          |||: :||
Db       3134 SRTAFAHRAAVTAADRDGFLDGLATLAQGGTSAHVHLDTARDGT--------TAFLFTGQ 3185

Qy       2644 GTQRPGMAHGLYHTHPVFAAALNDICTHLDPHLDHPLLPLLTQNDNDNEDAAALLQQTRY 2703
             |:||||    ||  ||||| ||::|| ||| ||: ||| ::   :    | |||| :|||
Db       3186 GSQRPGAGRELYDRHPVFARALDEICAHLDGHLELPLLDVMFAAEGSAE--AALLDETRY 3243

Qy       2704 AQPALFAFQVALHRLLTDGYHITPHYYAGHSLGEITAAHLAGILTLTDATTLITQRATLM 2763
             | |||| :||| ||: : : :    |||:||| |||:||: :| ||  |:  |  ||
Db       3244 TQCALFALEVALFRLV-ESWGMRPAALLGHSVGEIAAAHVAGVFSLADAARLVAARGRLM 3302

Qy       2764 QTMPP-GTMTTLHTTPHHITHHLTAHE---NDLAIAAINTPTSLVISGTPHTVQHITTLC 2819
             | :|  | |  :       |   |      | :||:| | : :|:||     :
Db       3303 QELPAGGAMLAVQAAEDEIRVWLETEERYAGRLDVAAVNGPEAAVLSGDADAAREAEAYW 3362

Qy       2820 QQQGIKTKTLPTNHAFHSPHTNPILNQLHQHTQTLTYHPPHTPLITANT----PPDQLLT 2875
                | :|: |  :|||| | : :|:     :|: :  |   :::  |       || |
Db       3363 SGLGRRTRALRVSHAFHSAHMDGMLDGFRAVLETVEFRRPSLTVVSNVTGLAAGPDDLCD 3422

Qy       2876 PHYWTQQARNTVDYATTTQTLHQHGVTTYIELGPDNTLTTLTHHNLPNPPTTTLT---LT 2932
             | || :  | || :    :  |   || | :|||||  ||  :    | : |  :     :
Db       3423 PEYWVRHVRGTVRFLDGVRVLRDLGVRTCLELGPDGVLTAMAADGLADTPADSAAGSPVG 3482

Qy       2933 HPHHHPQTH---------LLTNL-----AKTTTTWHPHHYTH-HDNQPHTHT-------- 2969
              |   |           ||  |     ::| |       | |   |  |
Db       3483 SPAGSPADSAAGALRPRPLLVALLRRKRSETETVADALGRAHAHGTGPDWHAWFAGSGAH 3542

Qy       2970 HLDLPTYPPQHHHYWLESTQPGAG---NVSAAGLDPTEHPLLGATLELATDGGALLAGRL 3026
              :|||||  |:   |||::  | |    : :  ||   :|||||| : |    | || |||
Db       3543 RVDLPTYSFRRDRYWLDA--PAADTAVDTAGLGLGTADHPLLGAVVSLPDRDGLLLTGRL 3600

Qy       3027 SLRSHPWLADHAVGGTVLLSGATFLELALHAGTYVGCDRVDELTLHAPLVVPVDGGVSVQ 3086
             |||:|||||||||| |:||| ||  :||| ||     |   | |||| |||:|  ||| ::
Db       3601 SLRTHPWLADHAVLGSVLLPGAAMVELAAHAAESAGLRDVRELTLLEPLVLPEHGGVELR 3660

Qy       3087 VGVAAADGE--------GRRLVSVYARGGSACGGGGASGGVWTCHASGVLVEAAAGGVVV 3138
             | | |  ||         | | ||:::|   |       :|  |:|||:|:|        |
Db       3661 VTVGAPAGEPGGESAGDGARPVSLHSRLADA-----PAGTAWSCHATGLLATDRPELPVA 3715

Qy       3139 DGLAGVWPPRGAVAVDVDGVRDRLAGAGCVLGPVFSGLRAVWRDGGDLLAEVCLP----- 3193
                | :|||:||  | :||: :|| | |   ||:| || ||||  |:: |:: ||
Db       3716 PDRAAMWPPQGAEEVPLDGLYERLDGNGLAFGPLFQGLNAVWRYEGEVFADIALPATTNA 3775

Qy       3194 -----EEAWGDAAG--FGLHPALLDGVVQPLSVLLPGGTGFGEGAGFGEGVRVPAVWGGV 3246
                  | ||   :|:|||||| :  ::|   ||          | ||||  | ||
Db       3776 TAPATANGGGSAAAAPYGIHPALLDASLHAIAV---GGL-----VDEPELVRVPFHWSGV 3827

Qy       3247 SLHRAGVTGVRVRVSAVGRGGGREAVSVVVGDEAGVPVASVDRLELRPVDMGQLRAVSVS 3306
             ::| ||    |||:::     | :|||: : |  | |: ||:|| ||||   |  |  |
Db       3828 TVHAAGAAAARVRLAS----AGTDAVSLSLTDGEGRPLVSVERLTLRPVTADQAAASRVG 3883

Qy       3307 AGRRGSLYAVQW----------------AEVGPVPVCGQ---------------------A 3330
                  | :: | |                  ||  | |:
Db       3884 ----GLMHRVAWRPYALASSGEQDPHATSYGPTAVLGKDELKVAAALESAGVEVGLYPDL 3939

Qy       3331 WAWHEDVGESGGGPVPGVVVLRCPDAGAGGGGGGGGGGVGEVVGGVLGVVQGWLGLERF 3390
                |  :||  : | | |  |:    |      |  ||  ||    |  | ::| ||  |
Db       3940 AALSQDV--AAGAPAPRTVLAPLP-----AGPADGGAEGVRGTVARTLELLQAWLADEHL 3992

Qy       3391 AGSRLVVVTRGAV----VAGPEDGPVDVVGASVWGLVRSAQAEHPDRFVLLDLDTDTGT- 3445
             ||:||::|||||     :| :||  |:  |: |||||:|| |:| || ||||  |  :
Db       3993 AGTRLLLVTRGAVRDPEGSGADDGGEDLSHAAAWGLVRTAQTENPGRFGLLDLADDASSY 4052

Qy       3446 ----DLDTGAGAGWGVDGGRVAAVVACGEPQLAVRGERLLAARLKRLESSGDVPAQRSGD 3501
                  : : ||                |||||:    :  |||
Db       4053 RTLPSVLSDAGLR-------------DEPQLALHDGTIRLARL---------------- 4082

Qy       3502 TRARRSDVPAQRSGGVPARRSVDVSGREVLPWLSGGSVLVTGGTGVLGAAVARHLAGVCG 3561
                  : |  :    ||          : |    |:||:||||| ||  ||||: |  |
Db       4083 -----ASVRPETGTAAPA----------LAP---EGTVLLTGGTGGLGGLVARHVVGEWG 4124

Qy       3562 VRDLLLVSRRGPDAPGAEGLRAELAALGAEVRIVACDVGERREVVRLLEGVPAGCPLTGV 3621
             || |||||||| |||||: |  || ||||:| : |||| :|  :  :|: :||  ||| |
Db       4125 VRRLLLVSRRGTDAPGADELVHELEALGADVSVAACDVADREALTAVLDAIPAEHPLTAV 4184

Qy       3622 VHAAGVLDDATIASLTPERLGTVFAAKVDAALLLDELTR--GMELSAFVLFSSAAGILGS 3679
             || |||| | |: |:| | :  |   ||||| ||||||      :|:|||:||||| : |
Db       4185 VHTAGVLSDGTLPSMTTEDVEHVLRPKVDAAFLLDELTSTPAYDLAAFVMFSSAAAVFGG 4244

Qy       3680 AGQGNYAAANAALDALAYRRRAAGLPGVSLAWGLWEEASGMTGHLAGTDHRRIIRSGLHP 3739
             |||| |||||| |||||:|||||||| :|| |||| | ||||| |    |. ||: |:|:
Db       4245 AGQGAYAAANATLDALAWRRRAAGLPALSLGWGLWAETSGMTGELGQADLRRMSRAGIGG 4304

Qy       3740 MSTPDALALFDAALALDR-PVLLP-----ADLRPA-----PPLPPLLQDLLPATRRR--- 3785
             :|  : :|| |||| || |||||      | || |       :| | :|:: |   |
Db       4305 ISDAEGIALLDAALRDDRHPVLLPLRLDAAGLRDAAGNDPAGIPALFRDVVGARTVRARP 4364

Qy       3786 ------TTRTT--TTGGADNGAQLHA-----RLAGQTHEQQHTTLLALVRSHIATVLGHT 3832
                   ||  |  | | ||  |:  |     | |       :   ||  |   :| ||||
Db       4365 SAASASTTAGTAGTPGTADGAAETAAVTLADRAATVDGPARQRLLLEFVVGEVAEVLGHA 4424

Qy       3833 TPDTIPPDRAFRDLGFDSLTAVELRNRLSRTTGLRLPTTLAFDHPNPTTLTHHLHTQL-- 3890
                  | :| | ||||||||||||||||:   || || || ||||:|  |  ||  :|
Db       4425 RGHRIDAERGFLDLGFDSLTAVELRNRLNSAGGLALPATLVFDHPSPAALASHLDAELPR 4484

Qy       3891 -QPQPDNA-----------------VAPVLAELDKLESALSALDKTDS-ASERVTLRLK 3930
                  | |                 :||:| :|| ||     :|:  || |   |:
Db       4485 GASDQDGAGNRNGNENGTTASRSTAETDALLAQLTRLEGALVLTGLSDAPGSEEVLEHLR 4544

Qy       3931 SLMLR-----------------------WNAPQHPTAESADDD--EKFTSATEAEIFK 3963
                ||                      | |       | |       | :|:  |:|
Db       4545 SLRSMVTGETGTGTASGAPDGAGSGAEDRPWAAGDGAGGGSEDGAGVPDFMNASAEELFG 4604

Qy       3964 FIDND 3968
                :| |
Db       4605 LLDQD 4609
```

RESULT 11
AAY77200
ID AAY77200 standard; Protein; 4613 AA.
XX
AC AAY77200;
XX
DT 05-JUN-2000 (first entry)
XX
DE S. venezuelae pik (macrolide biosynthesis) gene cluster protein #1.
XX
KW Desosamine biosynthesis; macrolide; polyketide; methymycin; pikromycin;
KW neomethymycin; narbomycin; polyhydroxyalkanoate monomer synthase;
KW biopolymer; antibiotic; chemotherapeutic; immunosuppressant; asthma;
KW chronic obstructive pulmonary disease; respiratory inflammation;
KW hypercholesterolaemia; crop protection agent.

XX
OS Streptomyces venezuelae ATCC15439.
XX
PN WO200000620-A2.
XX
PD 06-JAN-2000.
XX
PF 25-JUN-1999; 99WO-US14398.
XX
PR 26-JUN-1998; 98US-0105537.
XX
PA (MINU) UNIV MINNESOTA.
XX
PI Sherman DH, Liu H, Xue Y, Zhao L;.
XX
DR WPI; 2000-160679/14.
DR N-PSDB; AAZ87318.
XX
PT Desosamine and macrolide biosynthetic gene clusters, useful for, e .g.
PT synthesis of methymycin and pikromycin -
XX
PS Disclosure; Figure 31; 438pp; English.
XX
CC The invention relates to an isolated and purified nucleic acid segment
CC comprising a desosamine biosynthetic gene cluster, a fragment or its
CC biologically active variant, where the nucleic acid sequence is not
CC derived from the eryC gene cluster of Saccharopolyspora erythraea or
CC Streptomyces antibioticus. The invention also relates to a macrolide
CC biosynthetic gene cluster, or fragments thereof. The macrolide
CC biosynthetic gene cluster encodes proteins which synthesise methymycin,
CC pikromycin, neomethymycin, narbomycin or a combination of these
CC compounds. Recombinant or augmented cells comprising the desosamine
CC and/or macrolide biosynthetic gene clusters are useful for the
CC production of biologically active macrolides. The macrolide biosynthetic
CC proteins are useful for synthesis of methymycin, pikromycin,
CC neomethymycin and narbomycin. The alternative termination of polyketide
CC synthesis may be useful to prepare novel antibiotics and
CC polyhydroxyalkanoate (PHA) monomers. The compounds produced by the
CC recombinant host cells are useful as biopolymers, e.g., in packaging or
CC biomedical applications, to engineer PHA monomer synthases or to prepare
CC biologically active agents, such as chemotherapeutics,
CC immunosuppressants, agents to treat asthma, chronic obstructive pulmonary
CC disease as well as other diseases involving respiratory inflammation,
CC cholesterol-lowering agents or macrolide-based antibiotics which are
CC active against a variety of organisms, e.g., bacteria, including
CC multi-drug resistant pneumococci and other respiratory pathogens, as well
CC as viral parasitic pathogens, or as crop protection agents (e.g.,
CC fungicides or insecticides) via expression of polyketides in plants.
CC Sequences AAY77200-Y77203 and AAY80997 represent proteins encoded by the
CC macrolide biosynthetic gene cluster (pik) from Streptomyces venezuelae
CC ATCC 15439, as given in figure 31.
XX
SQ Sequence 4613 AA;

Query Match 39.1%; Score 8148; DB 21; Length 4613;
Best Local Similarity 44.8%; Pred. No. 0;
Matches 1909; Conservative 463; Mismatches 1307; Indels 586; Gaps 94;

```
Qy        15 GEVLGVADEADGGVVFVFPGQGPQWPGMGRELLDASDVFRESVRACEAAFAPYVDWSVEQ 74
Db       620 GLVRGVASGV-GRVAFVFPGQGTQWAGMGAELLDSSAVFAAAMAECEAALSPYVDWSLEA 678

Qy        75 VLRDSPDAPGLDRVDVVQPTLFAVMISLAALWRSQGVEPCAVLGHSLGEIAAAHVSGGLS 134
Db       679 VVRQAPGAPTLERVDVVQPVTFAVMVSLARVWQHHGVTPQAVVGHSQGEIAAAYVAGALS 738

Qy       135 LADAARVVTLWSQA-QTTLAGTGALVSVAATPDELLPRIAPWTEDNPARLAVAAVNGPRS 193
Db       739 LDDAARVVTLRSKSIAAHLAGKGGMLSLALSEDAVLERLAGFD-----GLSVAAVNGPTA 793

Qy       194 TVVSGAREAVADLVADLTAAQVRTRMIPVDVPAHSPLMYAIEERVVSGLLPITPRPSRIP 253
Db       794 TVVSGDPVQIEELARACEADGVRARVIPVDYASHSRQVEIIESELAEVLAGLSPQAPRVP 853

Qy       254 FHSSVTGGRLDTRELDAAYWYRNMSSTVRFEPAAR-LLLQQGPKTFVEMSPHPVLTMGLQ 312
Db       854 FFSTLEGAWITEPVLDGGYWYRNLRHRVGFAPAVETLATDEGFTHFVEVSAHPVLTMALP 913

Qy       313 ELAPDLGDTTGTADTVIMGTLRRGQGTLDHFLTSLAQ---------------LRGHGET 356
Db       914 ------GTVTGLA------TLRRDNGGQDRLVASLAEAWANGLAVDWSPLLPSATGHHSD 961

Qy       357 SATTVLS------ARLTALSPT----------------------QQQSLLLDLVRAHTM 387
Db       962 LPTYAFQTERHWLGEIEALAPAGEPAVQPAVLRTEAAEPAELDRDEQLRVILDKVRAQTA 1021

Qy       388 AVLNDDGNERTASDAGPSASFAHLGFDSVMGVELRNRLSKATGLRLPVTLIFDHTTPAAV 447
Db      1022 QVL----GYATGGQIEVDRTFREAGCTSLTGVDLRNRINAAFGVRMAPSMIFDFPTPEAL 1077

Qy       448 AARLRTAALGHLDEDTAPVPDSPSGHGGTAAADDPIAIIGMACRFPGGVRSPKDLWELAA 507
Db      1078 AEQLLLVVHGEAAANPAGAEPAPVAAAG--AVDEPVAIVGMACRLPGGVASPEDLWRLVA 1135

Qy       508 SGGDAIGPFPTDRGWPTEQRHAQDPTQPGTFYPQGGGFLHDAAHFDAGFFGISPREALAM 567
Db      1136 GGGDAISEFPQDRGWDVEGLYHPDPEHPGTSYVRQGGFIENVAGFDAAFFGISPREALAM 1195

Qy       568 DPQQRLLLETSWEAFERAGIDPLSVRGSRTGVFAGALSFDYGPRMDTASSEGAADVEGHI 627
Db      1196 DPQQRLLLETSWEAVEDAGIDPTSLRGRQVGVFTGAMTHEYGPSL----RDGGEGLDGYL 1251

Qy       628 LTGTTGSVLSGRIAYSFGLEGPAITVDTGCSASLVTLHLACQSLRSGECTLALAGGVSVM 687
Db      1252 LTGNTASVMSGRVSYTLGLEGPALTVDTACSSSLVALHLAVQALRKGEVDMALAGGVAVM 1311

Qy       688 STLGMFIEFSRQRGLSVDGRCKAYSAAADGTGWGEGVGMLLVERLSDAVRLGHRVLAVVR 747
Db      1312 PTPGMFVEFSRQRGLAGDGRSKAFAASADGTSWSEGVGVLLVERLSDARRNGHQVLAVVR 1371

Qy       748 GSAVNQDGASNGLTAPNGPAQERVIRQALANAGLSVADVDVVEGHGTGTTLGDPIEAQAL 807
Db      1372 GSALNQDGASNGLTAPNGPSQQRVIRRALADARLTTSDVDVVEAHGTGTRLGDPIEAQAL 1431

Qy       808 LATYGQ-RAGDRPLWLGSLKSNIGHTMAAAGVGGVIKMVMALREGVLPRTLHVDKPSPQV 866
Db      1432 IATYGQGRDDEQPLRLGSLKSNIGHTQAAAGVSGVIKMVQAMRHGLLPKTLHVDEPSDQI 1491

Qy       867 DWSAGAVRLLTEAVPWPGDAAGRLRRAGVSSFGIGGTNAHVILEEAPAAGGCVAGGGVLE 926
Db      1492 DWSAGAVELLTEAVDWPEKQDGGLRRAAVSSFGISGTNAHVVLEEAPV---------VVE 1542

Qy       927 GAPGLAISVAESVAAPVAVSAPVAESVPVPVPVPVPVPVSARSEAGLRAQAEALRQYVAV 986
Db      1543 GA-----SVVEPSVGGSAVGGGV----------TPWVVSAKSAAALDAQIERLAAF-AS 1585

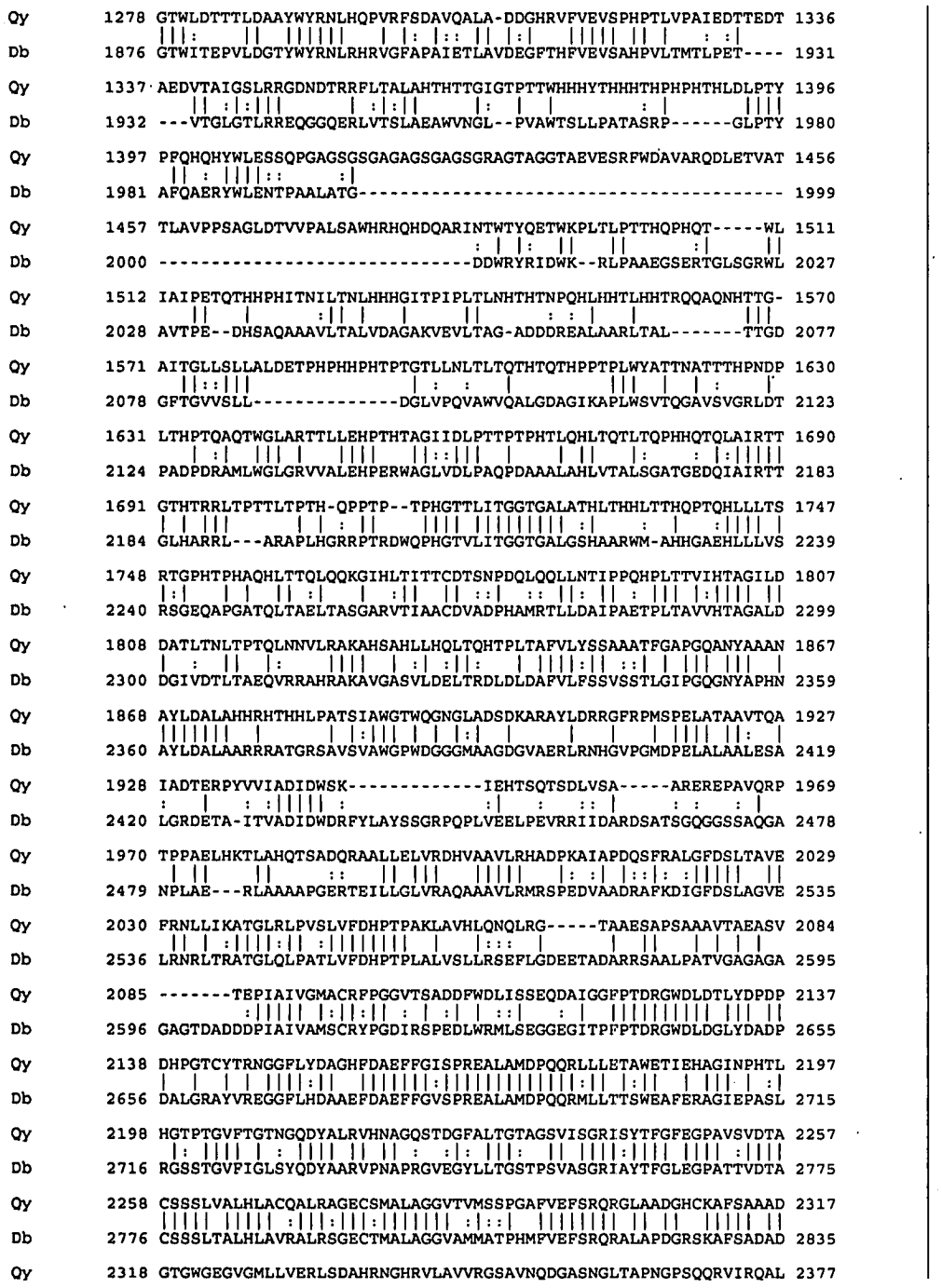
Qy       987 RPDVSLADVGA--------GLACGRAVLEHRAVVLAADREELVQGLGALAAGEPDRRVTT 1038
Db      1586 RDRTDDADAGAVDAGAVAHVLADGRAQFEHRAVALGAGADDLVQAL-----ADPDGLI-R 1639

Qy      1039 GHAPGGDRGGVVFVFPGQGGQWAGMGVRLLASSPVFARRMQACEEALAPWVDWSVVDILR 1098
Db      1640 GTASG--VGRVAFVFPGQGTQWAGMGAELLDSSAVFAAAMAECEAALSPYVDWSLEAVVR 1697

Qy      1099 RDAGDAVWERADVVQPVLFSVMVSLAALWRSYGIEPDAVLGHSQGEIAAAHVCGALSLKD 1158
Db      1698 QAPGAPTLERVDVVQPVTFAVMVSLARVWQHHGVTPQAVVGHSQGEIAAAYVAGALPLDD 1757

Qy      1159 AAKTVALRSRALAA-VRGRGGMASVPLPAQEVEQLIGERWAGRLWVAAVNGPRSTAVSGD 1217
Db      1758 AARVVTLRSKSIAAHLAGKGGMLSLALNEDAVLERLSD--FDGLSVAAVNGPTATVVSGD 1815

Qy      1218 AEAVDEVLAYCAGTGVRARRIPVDYASHCPHVQPLREELLELLGDISPQPSGVPFFSTVE 1277
Db      1816 PVQIEELAQACKADGFRARIIPVDYASHSRQVEIIESELAQVLAGLSPQAPRVPFFSTLE 1875
```


Qy 1278 GTWLDTTTLDAAYWYRNLHQPVRFSDAVQALA-DDGHRVFVEVSPHPTLVPAIEDTTEDT 1336
Db 1876 GTWITEPVLDGTYWYRNLRHRVGFAPAIETLAVDEGFTHFVEVSAHPVLTMTLPET---- 1931
Qy 1337 AEDVTAIGSLRRGDNDTRRFLTALAHTHTTGIGTPTTWHHHYTHHHTHPHPHTHLDLPTY 1396
Db 1932 ---VTGLGTLRREQGGQERLVTSLAEAWVNGL--PVAWTSLLPATASRP------GLPTY 1980
Qy 1397 PFQHQHYWLESSQPGAGSGSGAGAGSGAGSGRAGTAGGTAEVESRFWDAVARQDLETVAT 1456
Db 1981 AFQAERYWLENTPAALATG---------- 1999
Qy 1457 TLAVPPSAGLDTVVPALSAWHRHQHDQARINTWTYQETWKPLTLPTTHQPHQT-----WL 1511
Db 2000 ----------DDWRYRIDWK--RLPAAEGSERTGLSGRWL 2027
Qy 1512 IAIPETQTHHPHITNILTNLHHHGITPIPLTLNHTHTNPQHLHHTLHHTRQQAQNHTTG- 1570
Db 2028 AVTPE--DHSAQAAAVLTALVDAGAKVEVLTAG-ADDDREALAARLTAL-------TTGD 2077
Qy 1571 AITGLLSLLALDETPHPHHPHTPTGTLLNLTLTQTHTQTHPPTPLWYATTNATTTHPNDP 1630
Db 2078 GFTGVVSLL----------DGLVPQVAWVQALGDAGIKAPLWSVTQGAVSVGRLDT 2123
Qy 1631 LTHPTQAQTWGLARTTLLEHPTHTAGIIDLPTTPTPHTLQHLTQTLTQPHHQTQLAIRTT 1690
Db 2124 PADPDRAMLWGLGRVVALEHPERWAGLVDLPAQPDAAALAHLVTALSGATGEDQIAIRTT 2183
Qy 1691 GTHTRRLTPTTLTPTH-QPPTP--TPHGTTLITGGTGALATHLTHHLTTHQPTQHLLLTS 1747
Db 2184 GLHARRL---ARAPLHGRRPTRDWQPHGTVLITGGTGALGSHAARWM-AHHGAEHLLLVS 2239
Qy 1748 RTGPHTPHAQHLTTQLQQKGIHLTITTCDTSNPDQLQQLLNTIPPQHPLTTVIHTAGILD 1807
Db 2240 RSGEQAPGATQLTAELTASGARVTIAACDVADPHAMRTLLDAIPAETPLTAVVHTAGALD 2299
Qy 1808 DATLTNLTPTQLNNVLRAKAHSAHLLHQLTQHTPLTAFVLYSSAAATFGAPGQANYAAAN 1867
Db 2300 DGIVDTLTAEQVRRAHRAKAVGASVLDELTRDLDLDAFVLFSSVSSTLGIPGQGNYAPHN 2359
Qy 1868 AYLDALAHHRHTHHLPATSIAWGTWQGNGLADSDKARAYLDRRGFRPMSPELATAAVTQA 1927
Db 2360 AYLDALAARRRATGRSAVSVAWGPWDGGGMAAGDGVAERLRNHGVPGMDPELALAALESA 2419
Qy 1928 IADTERPYVVIADIDWSK----------IEHTSQTSDLVSA-----ARE-REPAVQRP 1969
Db 2420 LGRDETA-ITVADIDWDRFYLAYSSGRPQPLVEELPEVRRIIDARDSATSGQGGSSAQGA 2478
Qy 1970 TPPAELHKTLAHQTSADQRAALLELVRDHVAAVLRHADPKAIAPDQSFRALGFDSLTAVE 2029
Db 2479 NPLAE---RLAAAAPGERTEILLGLVRAQAAAVLRMRSPEDVAADRAFKDIGFDSLAGVE 2535
Qy 2030 FRNLLIKATGLRLPVSLVFDHPTPAKLAVHLQNQLRG-----TAAESAPSAAAVTAEASV 2084
Db 2536 LRNRLTRATGLQLPATLVFDHPTPLALVSLLRSEFLGDEETADARRSAALPATVGAGAGA 2595
Qy 2085 -------TEPIAIVGMACRFPGGVTSADDFWDLISSEQDAIGGFPTDRGWDLDTLYDPDP 2137
Db 2596 GAGTDADDDPIAIVAMSCRYPGDIRSPEDLWRMLSEGGEGITPFPTDRGWDLDGLYDADP 2655
Qy 2138 DHPGTCYTRNGGFLYDAGHFDAEFFGISPREALAMDPQQRLLLETAWETIEHAGINPHTL 2197
Db 2656 DALGRAYVREGGFLHDAAEFDAEFFGVSPREALAMDPQQRMLLTTSWEAFERAGIEPASL 2715
Qy 2198 HGTPTGVFTGTNGQDYALRVHNAGQSTDGFALTGTAGSVISGRISYTFGFEGPAVSVDTA 2257
Db 2716 RGSSTGVFIGLSYQDYAARVPNAPRGVEGYLLTGSTPSVASGRIAYTFGLEGPATTVDTA 2775
Qy 2258 CSSSLVALHLACQALRAGECSMALAGGVTVMSSPGAFVEFSRQRGLAADGHCKAFSAAAD 2317
Db 2776 CSSSLTALHLAVRALRSGECTMALAGGVAMMATPHMFVEFSRQRALAPDGRSKAFSADAD 2835
Qy 2318 GTGWGEGVGMLLVERLSDAHRNGHRVLAVVRGSAVNQDGASNGLTAPNGPSQQRVIRQAL 2377
Db 2836 GFGAAEGVGLLLVERLSDARRNGHPVLAVVRGTAVNQDGASNGLTAPNGPSQQRVIRQAL 2895
Qy 2378 ANAGLSAGDVDAVEAHGTGTTLGDPIEAQALLATYGQDRAGEGPLWLGSVKSNVGHTQAA 2437
Db 2896 ADARLAPGDIDAVETHGTGTSLGDPIEAQGLQATYGKERPAERPLAIGSVKSNIGHTQAA 2955
Qy 2438 AGVAGVIKMVMALRHGLLPRTLHVDEPSPHVDWSAGAVQLLTETVPWPGGEGRLRRAGVS 2497
Db 2956 AGAAGIIKMVLAMRHGTLPKTLHADEPSPHVDWANSGLALVTEPIDWPAGTGP-RRAAVS 3014
Qy 2498 SFGVSGTNAHVILEEAP--------ADDVP----------GGPPAGEGDAGSDDEAAA 2537
Db 3015 SFGISGTNAHVVLEQAPDAAGEVLGADEVPEVSETVAMAGTAGTSEVAEGSEASEAPAAP 3074
Qy 2538 GS-----PGVWPWLVSAKSQPALRAQAQALHAHLTDHPGLDLA---------DVGYTLAH 2583
Db 3075 GSREASLPGHLPWVLSAKDEQSLRGQAAALHAWLSE-PAADLSDADGPARLRDVGYTLAT 3133
Qy 2584 ARAVFDHRATLIAADRDTFLQALQALAAGEPHPAVIHSSAPGGTGTGEAAGKTAFICSGQ 2643
Db 3134 SRTAFAHRAAVTAADRDGFLDGLATLAQGGTSAHVHLDTARDGT--------TAFLFTGQ 3185
Qy 2644 GTQRPGMAHGLYHTHPVFAAALNDICTHLDPHLDHPLLPLLTQNDNDNEDAAALLQQTRY 2703
Db 3186 GSQRPGAGRELYDRHPVFARALDEICAHLDGHLELPLLDVMFAAEGSAE--AALLDETRY 3243
Qy 2704 AQPALFAFQVALHRLLTDGYHITPHYYAGHSLGEITAAHLAGILTLTLTDATTLITQRATLM 2763
Db 3244 TQCALFALEVALFRLV-ESWGMRPAALLGHSVGEIAAAHVAGVFSLADAARLVAARGRLM 3302
Qy 2764 QTMPP-GTMTTLHTTPHHITHHLTAHE---NDLAIAAINTPTSLVISGTPHTVQHITTLC 2819
Db 3303 QELPAGGAMLAVQAAEDEIRVWLETEERYAGRLDVAAVNGPEAAVLSGDADAAREAEAYW 3362
Qy 2820 QQQGIKTKTLPTNHAFHSPHTNPILNQLHQHTQTLTYHPPHTPLITANT----PPDQLLT 2875
Db 3363 SGLGRRTRALRVSHAFHSAHMDGMLDGFRAVLETVEFRRPSLTVVSNVTGLAAGPDDLCD 3422
Qy 2876 PHYWTQQARNTVDYATTTQTLHQHGVTTYIELGPDNTLTTLTHHNLPNPPTTTLT---LT 2932
Db 3423 PEYWVRHVRGTVRFLDGVRVLRDLGVRTCLELGPDGVLTAMAADGLADTPADSAAGSPVG 3482
Qy 2933 HPHHHPQTH---------LLTNL-----AKTTTTWHPHHYTH-HDNQPHTHT-------- 2969
Db 3483 SPAGSPADSAAGALRPRPLLVALLRRKRSETETVADALGRAHAHGTGPDWHAWFAGSGAH 3542
Qy 2970 HLDLPTYPFQHHHYWLESTQPGAG---NVSAAGLDPTEHPLLGATLELATDGGALLAGRL 3026
Db 3543 RVDLPTYSFRRDRYWLDA--PAADTAVDTAGLGLGTADHPLLGAVVSLPDRDGLLLTGRL 3600
Qy 3027 SLRSHPWLADHAVGGTVLLSGATFLELALHAGTYVGCDRVDELTLHAPLVVPVDGGVSVQ 3086
Db 3601 SLRTHPWLADHAVLGSVLLPGAAMVELAAHAAESAGLRDVRELTLLEPLVLPEHGGVELR 3660
Qy 3087 VGVAAADGE--------GRRLVSVYARGGSACGGGGASGGVWTCHASGVLVEAAAGGVVV 3138
Db 3661 VTVGAPAGEPGGESAGDGARPVSLHSRLADA-----PAGTAWSCHATGLLATDRPELPVA 3715
Qy 3139 DGLAGVWPPRGAVAVDVDGVRDRLAGAGCVLGPVFSGLRAVWRDGGDLLAEVCLP----- 3193
Db 3716 PDRAAMWPPQGAEEVPLDGLYERLDGNGLAFGPLFQGLNAVWRYEGEVFADIALPATTNA 3775
Qy 3194 -----EEAWGDAAG--FGLHPALLDGVVQPLSVLLPGGTGFGEGAGFGEGVRVPAVWGGV 3246
Db 3776 TAPATANGGGSAAAAPYGIHPALLDASLHAIAV---GGL-----VDEPELVRVPFHWSGV 3827
Qy 3247 SLHRAGVTGVRVRVSAVGRGGGREAVSVVVGDEAGVPVASVDRLELRPVDMGQLRAVSVS 3306
Db 3828 TVHAAGAAAARVRLAS----AGTDAVSLSLTDGEGRPLVSVERLTLRPVTADQAAASRVG 3883
Qy 3307 AGRRGSLYAVQW----------AEVGPVPVCGQ----------A 3330

```
Db       3884 ----GLMHRVAWRPYALASSGEQDPHATSYGPTAVLGKDELKVAAALESAGVEVGLYPDL 3939

Qy       3331 WAWHEDVGESGGGPVPGVVVLRCPDAGAGGGGGGGGGGGVGEVVGGVLGVVQGWLGLERF 3390

Db       3940 AALSQDV--AAGAPAPRTVLAPLP-----AGPADGGAEGVRGTVARTLELLQAWLADEHL 3992

Qy       3391 AGSRLVVVTRGAV----VAGPEDGPVDVVGASVWGLVRSAQAEHPDRFVLLDLDTDTGT- 3445

Db       3993 AGTRLLLVTRGAVRDPEGSGADDGGEDLSHAAAWGLVRTAQTENPGRFGLLDLADDASSY 4052

Qy       3446 ----DLDTGAGAGWGVDGGRVAAVVACGEPQLAVRGERLLAARLKRLESSGDVPAQRSGD 3501

Db       4053 RTLPSVLSDAGLR--------------DEPQLALHDGTIRLARL---------------- 4082

Qy       3502 TRARRSDVPAQRSGGVPARRSVDVSGREVLPWLSGGSVLVTGGTGVLGAAVARHLAGVCG 3561

Db       4083 -----ASVRPETGTAAPA----------LAP---EGTVLLTGGTGGLGGLVARHVVGEWG 4124

Qy       3562 VRDLLLVSRRGPDAPGAEGLRAELAALGAEVRIVACDVGERREVVRLLEGVPAGCPLTGV 3621

Db       4125 VRRLLLVSRRGTDAPGADELVHELEALGADVSVAACDVADREALTAVLDAIPAEHPLTAV 4184

Qy       3622 VHAAGVLDDATIASLTPERLGTVFAAKVDAALLLDELTR--GMELSAFVLFSSAAGILGS 3679

Db       4185 VHTAGVLSDGTLPSMTTEDVEHVLRPKVDAAFLLDELTSTPAYDLAAFVMFSSAAAVFGG 4244

Qy       3680 AGQGNYAAANAALDALAYRRRAAGLPGVSLAWGLWEEASGMTGHLAGTDHRRIIRSGLHP 3739

Db       4245 AGQGAYAAANATLDALAWRRRAAGLPALSLGWGLWAETSGMTGELGQADLRRMSRAGIGG 4304

Qy       3740 MSTPDALALFDAALALDR-PVLLP-----ADLRPA-----PPLPPLLQDLLPATRRR--- 3785

Db       4305 ISDAEGIALLDAALRDDRHPVLLPLRLDAAGLRDAAGNDPAGIPALFRDVVGARTVRARP 4364

Qy       3786 ------TTRTT--TTGGADNGAQLHA-----RLAGQTHEQQHTTLLALVRSHIATVLGHT 3832

Db       4365 SAASASTTAGTAGTPGTADGAAETAAVTLADRAATVDGPARQRLLLEFVVGEVAEVLGHA 4424

Qy       3833 TPDTIPPDRAFRDLGFDSLTAVELRNRLSRTTGLRLPTTLAFDHPNPTTLTHHLHTQL-- 3890

Db       4425 RGHRIDAERGFLDLGFDSLTAVELRNRLNSAGGLALPATLVFDHPSPAALASHLDAELPR 4484

Qy       3891 -QPQPDNA-----------------VAPVLAELDKLESALSALDKTDS-ASERVTLRLK 3930

Db       4485 GASDQDGAGNRNGNENGTTASRSTAETDALLAQLTRLEGALVLTGLSDAPGSEEVLEHLR 4544

Qy       3931 SLMLR------------------------WNAPQHPTAESADDD--EKFTSATEAEIFK 3963

Db       4545 SLRSMVTGETGTGTASGAPDGAGSGAEDRPWAAGDGAGGGSEDGAGVPDFMNASAEELFG 4604

Qy       3964 FIDND 3968

Db       4605 LLDQD 4609


RESULT 12
AAY77180
ID   AAY77180 standard; Protein; 12199 AA.
XX
AC   AAY77180;
XX
DT   05-JUN-2000  (first entry)
XX
DE   S. venezuelae pik gene cluster-encoded protein, SEQ ID NO:6.
XX
KW   Desosamine biosynthesis; macrolide; polyketide; methymycin; pikromycin;
KW   neomethymycin; narbomycin; polyhydroxyalkanoate monomer synthase;
KW   biopolymer; antibiotic; chemotherapeutic; immunosuppressant; asthma;
KW   chronic obstructive pulmonary disease; respiratory inflammation;
KW   hypercholesterolaemia; crop protection agent.
XX
OS   Streptomyces venezuelae ATCC15439.
XX
PN   WO200000620-A2.
XX
PD   06-JAN-2000.
XX
PF   25-JUN-1999;   99WO-US14398.
XX
PR   26-JUN-1998;   98US-0105537.
XX
PA   (MINU ) UNIV MINNESOTA.
XX
PI   Sherman DH,  Liu H,  Xue Y,  Zhao L;
XX
DR   WPI; 2000-160679/14.
DR   N-PSDB; AAZ87285.
XX
PT   Desosamine and macrolide biosynthetic gene clusters, useful for, e .g.
PT   synthesis of methymycin and pikromycin -
XX
PS   Claim 19; Page 315-353; 438pp; English.
XX
CC   The invention relates to an isolated and purified nucleic acid segment
CC   comprising a desosamine biosynthetic gene cluster, a fragment or its
CC   biologically active variant, where the nucleic acid sequence is not
CC   derived from the eryC gene cluster of Saccharopolyspora erythraea or
CC   Streptomyces antibioticus. The invention also relates to a macrolide
CC   biosynthetic gene cluster, or fragments thereof. The macrolide
CC   biosynthetic gene cluster encodes proteins which synthesise methymycin,
CC   pikromycin, neomethymycin, narbomycin or a combination of these
CC   compounds. Recombinant or augmented cells comprising the desosamine
CC   and/or macrolide biosynthetic gene clusters are useful for the production
CC   of biologically active macrolides. The macrolide biosynthetic proteins
CC   are useful for synthesis of methymycin, pikromycin, neomethymycin and
CC   narbomycin. The alternative termination of polyketide synthesis may be
CC   useful to prepare novel antibiotics and polyhydroxyalkanoate (PHA)
CC   monomers. The compounds produced by the recombinant host cells are useful
CC   as biopolymers, e.g., in packaging or biomedical applications, to
CC   engineer PHA monomer synthases or to prepare biologically active agents,
CC   such as chemotherapeutics, immunosuppressants, agents to treat asthma,
CC   chronic obstructive pulmonary disease as well as other diseases involving
CC   respiratory inflammation, cholesterol-lowering agents or macrolide-based
CC   antibiotics which are active against a variety of organisms, e.g.,
CC   bacteria, including multi-drug resistant pneumococci and other
CC   respiratory pathogens, as well as viral parasitic pathogens, or as crop
CC   protection agents (e.g., fungicides or insecticides) via expression of
CC   polyketides in plants. The present sequence represents a protein
CC   encoded by the macrolide biosynthetic gene cluster (pik) from
CC   Streptomyces venezuelae ATCC 15439.
XX
SQ   Sequence   12199 AA;

  Query Match             39.1%;  Score 8148;  DB 21;  Length 12199;
  Best Local Similarity   44.8%;  Pred. No. 0;
  Matches 1909;  Conservative  463;  Mismatches 1307;  Indels  586;  Gaps   94;

Qy         15 GEVLGVADEADGGVVFVFPGQGPQWPGMGRELLDASDVFRESVRACEAAFAPYVDWSVEQ 74

Db       1278 GLVRGVASGV-GRVAFVFPGQGTQWAGMGAELLDSSAVFAAAMAECEAALSPYVDWSLEA 1336

Qy         75 VLRDSPDAPGLDRVDVVQPTLFAVMISLAALWRSQGVEPCAVLGHSLGEIAAAHVSGGLS 134

Db       1337 VVRQAPGAPTLERVDVVQPVTFAVMVSLARVWQHHGVTPQAVVGHSQGEIAAAYVAGALS 1396

Qy        135 LADAARVVTLWSQA-QTTLAGTGALVSVAATPDELLPRIAPWTEDNPARLAVAAVNGPRS 193

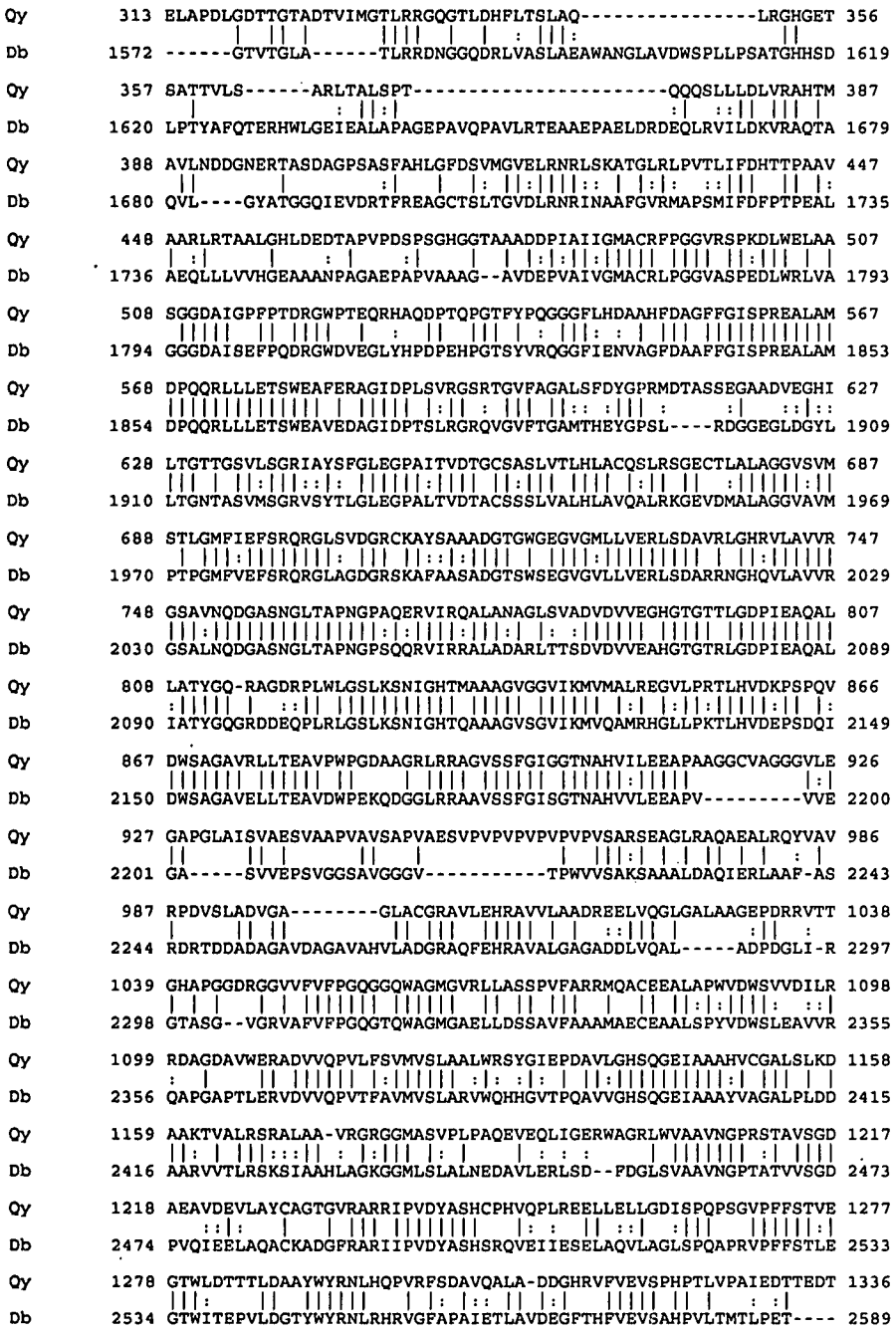
Db       1397 LDDAARVVTLRSKSIAAHLAGKGGMLSLALSEDAVLERLAGFD-----GLSVAAVNGPTA 1451

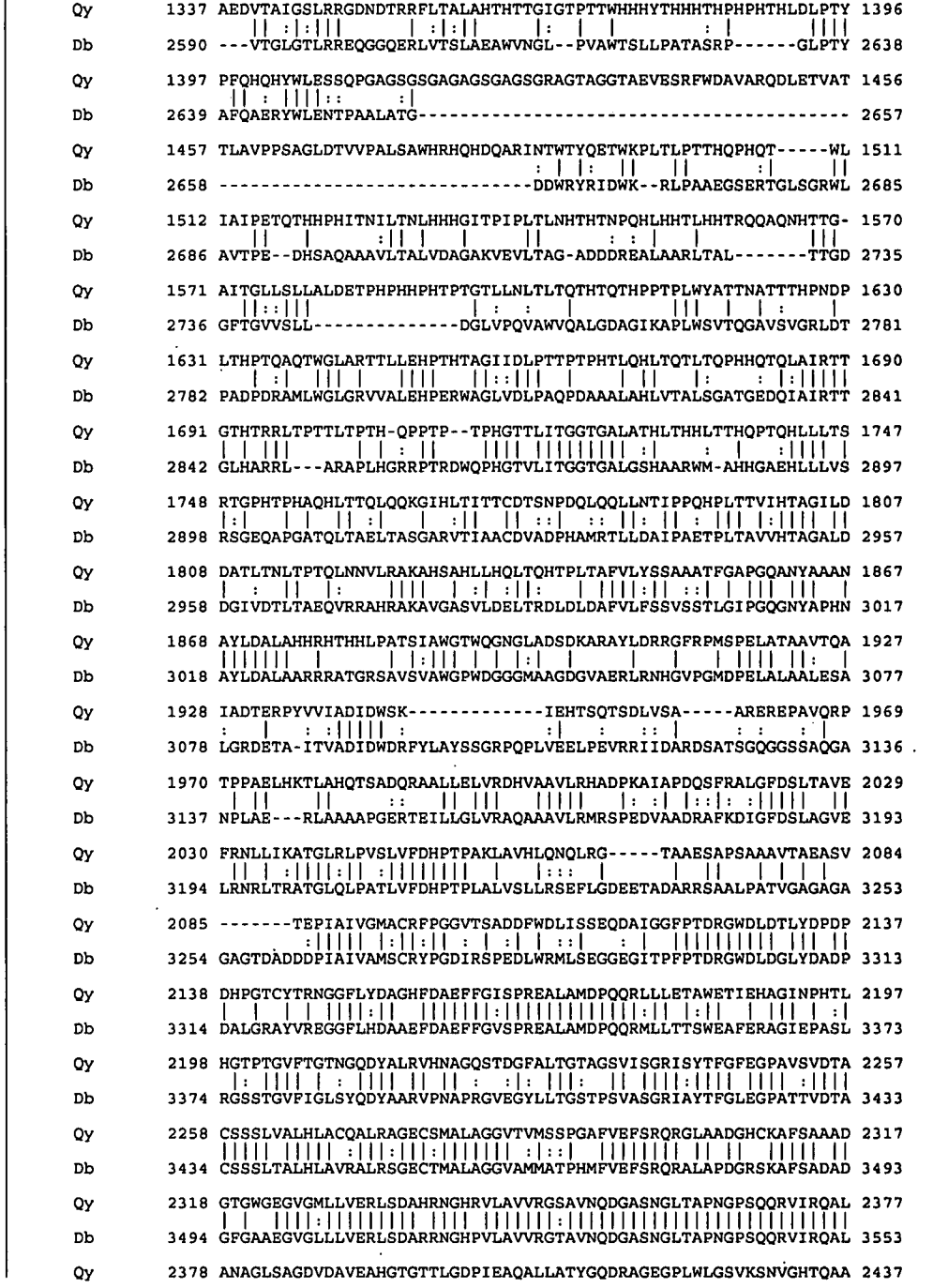
Qy        194 TVVSGAREAVADLVADLTAAQVRTRMIPVDVPAHSPLMYAIEERVVSGLLPITPRPSRIP 253

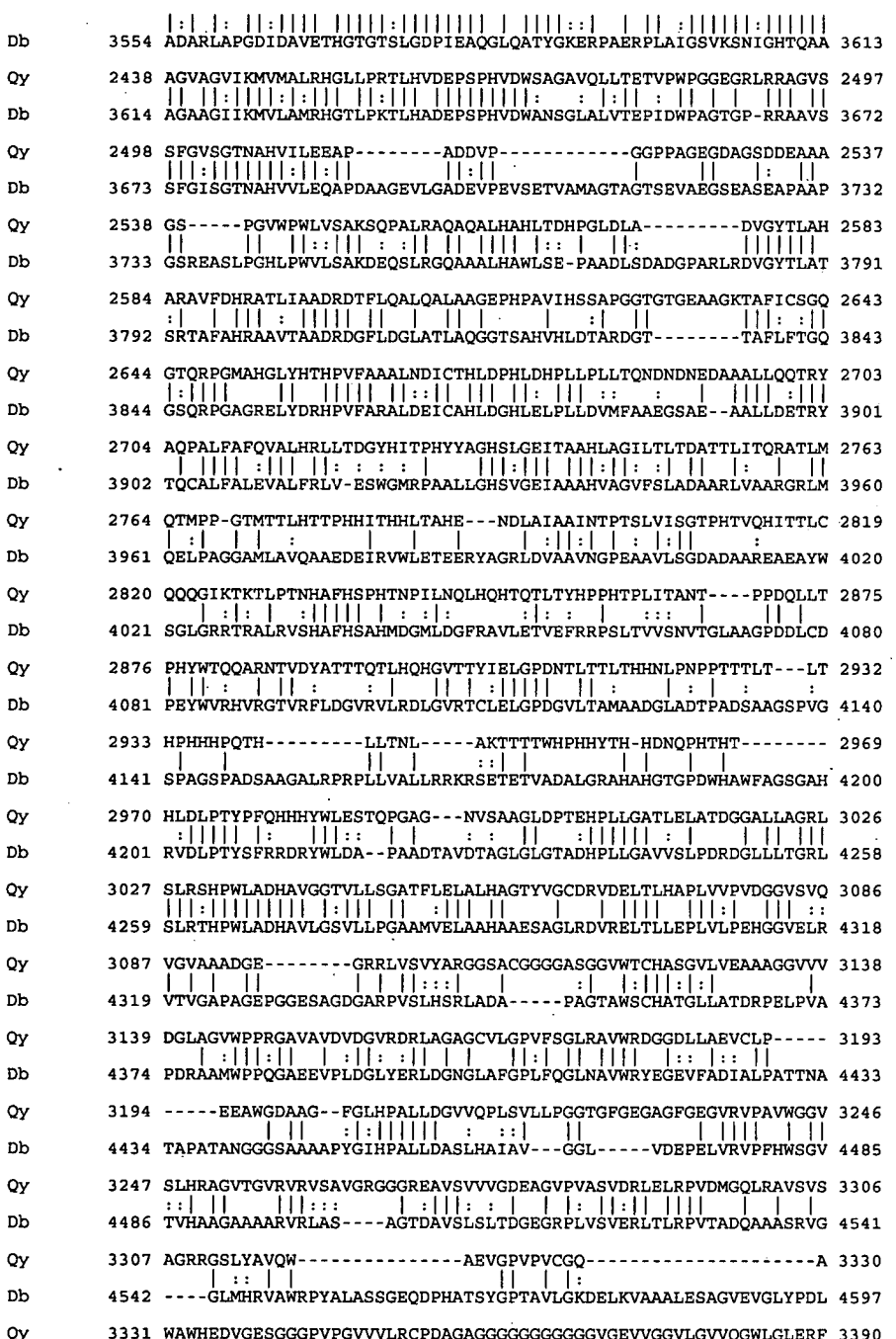
Db       1452 TVVSGDPVQIEELARACEADGVRARVIPVDYASHSRQVEIIESELAEVLAGLSPQAPRVP 1511

Qy        254 FHSSVTGGRLDTRELDAAYWYRNMSSTVRFEPAAR-LLLQQGPKTFVEMSPHPVLTMGLQ 312

Db       1512 FFSTLEGAWITEPVLDGGYWYRNLRHRVGFAPAVETLATDEGFTHFVEVSAHPVLTMALP 1571
```







RESULT 13
AAW22601
ID AAW22601 standard; Protein; 4472 AA.
XX
AC AAW22601;
XX
DT 27-FEB-1998 (first entry)
XX
DE Tylactone synthase ORF1 protein.
XX
KW Tylactone synthase gene cluster; tylG gene; multifunctional protein;
KW polyketide; tylactone synthesis; antibiotic; tylosin.
XX
OS Streptomyces fradiae.
XX
FH Key Location/Qualifiers
FT Misc-difference 1
FT /note= "encoded by GTG"
FT Domain 43..447
FT /note= "ketosynthase domain, KSq(s)"
FT Domain 586..914
FT /note= "acyltransferase domain, AT(s)"

```
FT   Domain          954..1038
FT                   /note= "acyl carrier protein domain, ACP(s)"
FT   Domain          1060..1483
FT                   /note= "ketosynthase domain, KS1"
FT   Domain          1606..1934
FT                   /note= "acyltransferase domain, AT1"
FT   Domain          2199..2378
FT                   /note= "ketoreductase domain, KR1"
FT   Domain          2469..2552
FT                   /note= "acyl carrier protein domain, ACP1"
FT   Domain          2576..2999
FT                   /note= "ketosynthase domain, KS2"
FT   Domain          3149..3477
FT                   /note= "acyltransferase domain, AT2"
FT   Domain          3502..3687
FT                   /note= "dehydratase domain, DH2"
FT   Domain          4016..4200
FT                   /note= "ketoreductase domain, KR2"
FT   Domain          4302..4385
FT                   /note= "acyl carrier protein domain, ACP2"
XX
PN   EP791655-A2.
XX
PD   27-AUG-1997.
XX
PF   19-FEB-1997;   97EP-0301056.
XX
PR   22-FEB-1996;   96US-0012078.
XX
PA   (ELIL ) LILLY & CO ELI.
XX
PI   Dehoff BS,  Kuhstoss SA,  Rosteck PR,  Sutton KL;
XX
DR   WPI; 1997-418046/39.
DR   N-PSDB; AAT80413.
XX
PT   DNA encoding Streptomyces fradiae tylactone synthase domain - for
PT   production of tylosin-related polyketide compounds
XX
PS   Claim 9; Pages 66-80; 220pp; English.
XX
CC   AAW22601-W22605 represent proteins encoded by the tylactone synthase
CC   gene cluster of the invention. The gene cluster is also referred to as
CC   the tylG gene, and was isolated from Streptomyces fradiae. These
CC   sequences are multifunctional proteins which direct the synthesis of the
CC   polyketide tylactone, isolated from Streptomyces fradiae. Tylactone is
CC   the basic building block of the antibiotic tylosin. The DNA sequence can
CC   be modified so as to alter the type of carboxylic acids incorporated,
CC   the number of carboxylic acids incorporated and/or the post-condensation
CC   reactions performed, thereby resulting in novel tylosin-related
CC   polyketides.
XX
SQ   Sequence   4472 AA;

  Query Match            37.0%;  Score 7705.5;  DB 18;  Length 4472;
  Best Local Similarity  44.3%;  Pred. No. 0;
  Matches 1849;  Conservative  453;  Mismatches 1380;  Indels  495;  Gaps   97;

Qy         13 AAGEVLGVADEADGGVVFVFPGQGPQWPGMGRELLDASDVFRESVRACEAAFAPYVDWSV  72
              |||           |||| |||||| || |||  || :|:||  |:| |  | : :| | :
Db        571 AAGGAGAAGGAGGGGVVLVFPGQGTQWVGMGAGLLGSSEVFAASMRECARALSVHVGWDL  630

Qy         73 EQVLRDSPDAPGLDRVDVVQPTLFAVMISLAALWRSQGVEPCAVLGHSLGEIAAAHVSGG  132
                :|:      ||:|||||||  :|||:|||  |:: ||:  ||:||| |||||| |:|
Db        631 LEVV---SGGAGLERVDVVQPVTWAVMVSLARYWQAMGVDVAAVVGHSQGEIAAATVAGA  687

Qy        133 LSLADAARVVTLWS-QAQTTLAGTGALVSVAATPDELLPRIAPWTEDNPARLAVAAVNGP  191
              ||| ||| || | :       ||| ||: :|     |:   :| |       : |||||||
Db        688 LSLEDAAAVVALRAGLIGRYLAGRGAMAAVPLPAGEVEAGLAKW-----PGVEVAAVNGP  742

Qy        192 RSTVVSGAREAVADLVADLTAAQVRTRMIPVDVPAHSPLMYAIE---ERVVSGLLPITPR  248
               |||||| | |||  ||    |  |: |:|||| :|| :  ::   |||:||   | ||
Db        743 ASTVVSGDRRAVAGYVAVCQAEGVQARLIPVDYASHSRHVEDLKGELERVLSG---IRPR  799

Qy        249 PSRIPFHSSVTGGRLDTRELDAAYWYRNMSSTVRFEPAARLLLQQGPKTFVEMSPHPVLT  308
                |:|  |:| | :       || ||:||: : | |       ||::| : |:|:| ||||
Db        800 SPRVPVCSTVAGEQPGEPVFDAGYWFRNLRNRVEPSAVVGGLLEEGHRRPIEVSAHPVLV  859

Qy        309 MGLQELAPDLGDTTGTADTVI--MGTLRRGQGTLDHFLTSLAQLRGHGET----------  356
                :::      |   ||  :  |||||   :    ||| |:   || |
Db        860 HAIEQ-------TAEAADRSVHATGTLRRQDDSPHRLLTSTAEAWAHGATLTWDPALPPG  912

Qy        357 SATTV-------------------------------LSARLTALSPTQQQ-SLLLDLVRA  384
                ||:                             |:  |  |  :|:|   |||||
Db        913 HLTTLPTYPFNHHHYWLDTIDGGGGDDATQEKESGPLTRELRGLPSSQKQLGFLLDLVCR  972

Qy        385 HTMAVLNDDGNERTASDAGPSASFAHLGFDSVMGVELRNRLSKATGLRLPVTLIFDHTTP  444
              ||  ||  |    ||::  |  |:|   |  |: ||||||||   ||| || ||::|  ||
Db        973 HTAVVLGLD----TAAEVDPDLSFKKQGIQSMTGVELRNRLLTETGLALPTTLVYDRPTP 1028

Qy        445 AAVAARLRTAALGHLDEDTAPVPDSPSGHGGTAA-----ADDPIAIIGMACRFPGGVRSP  499
               |:|  | | |          ||||     |     | :|||::|| ||||||| ||
Db       1029 RALAQFLHTELL----------DGSPSGSVLAPAQKSFEAQEPIAVVGMGCRFPGGVGSP 1078

Qy        500 KDLWELAASGGDAIGPFPTDRGWPTEQRHAQDPTQPGTFYPQGGGFLHDAAHFDAGFFGI  559
              : || |   | ||: ||| ||||  |  :  :|   |  | : |||||||| ||| ||||
Db       1079 EALWRLVVEGVDAVSPFPGDRGWDVEGLYDPEPGVAGKSYVREGGFLHDAAEFDAEFFGI 1138

Qy        560 SPREALAMDPQQRLLLETSWEAFERAGIDPLSVRGSRTGVFAGALSFDYGPRMDTASSEG  619
              |||||:|||||||||||||||| |||||| |: ||||||:|| :  :||||:    :||
Db       1139 SPREAVAMDPQQRLLLETSWEAIERAGIDPHSLHGSRTGVYAGVMPQEYGPRL----AEG 1194

Qy        620 AADVEGHILTGTTGSVLSGRIAYSPGLEGPAITVDTGCSASLVTLHLACQSLRSGECTLA  679
              |   :|::||||:|||:|||:||: ||||||:|||| ||:||| |||| |:|| ||| :|
Db       1195 AEGSDGYLLTGTSGSVVSGRVAYTLGLEGPAVTVDTACSSSLVALHLAVQALRGGECDMA 1254

Qy        680 LAGGVSVMSTLGMFIEFSRQRGLSVDGRCKAYSAAADGTGWGEGVGMLLVERLSDAVRLG  739
              |||||:||:  |||:|||||||: ||||||::  |||| | || |::|||||||| |||
Db       1255 LAGGVTVMAGPGMFVEFSRQRGLAADGRCKAFADGADGTAWAEGAGVVLVERLSDARRLG 1314

Qy        740 HRVLAVVRGSAVNQDGASNGLTAPNGPAQERVIRQALANAGLSVADVDVVEGHGTGTTLG  799
              | ||||| |||||||||||||||:||:|||||||| || |:|||||||| ||||| ||
Db       1315 HPVLAVVCGSAVNQDGASNGLTAPSGPSQERVIRQALGNARLTVADVDVVEAHGTGTRLG 1374

Qy        800 DPIEAQALLATYGQ-RAGDRPLWLGSLKSNIGHTMAAAGVGGVIKMVMALREGVLPRTLH  858
              |||||||| |||: | | ||:|||||||||||  ||||| ||||||:|:| | ||||||
Db       1375 DPIEAQALLGTYGRDRDGGRPVWLGSLKSNIGHAQAAAGVAGVIKMVLAMRYGWLPRTLH 1434

Qy        859 VDKPSPQVDWSAGAVRLLTEAVPWPGDAAGRLRRAGVSSFGIGGTNAHVILEEAPAAGGC  918
              ||:||  |||||| | ||||  |||    | ||| ||:||: |||||:|||
Db       1435 VDEPSRHVDWSAGGVWLLTEAREWPG--VDRPRRAAVSAFGVSGTNAHLILE-------- 1484

Qy        919 VAGGGVLEGAPGLAISVAESVAAPVAVSAPVAESVPV--PVPVPVPVPVSARSEAGLRAQ  976
                        |   ||: :|   | : |:||  |        ||| || ||   :|
Db       1485 ------------APDTAEAESATTPVRSEVSESAAVLDARSGVVPVVVSGRSRVVVREA 1531

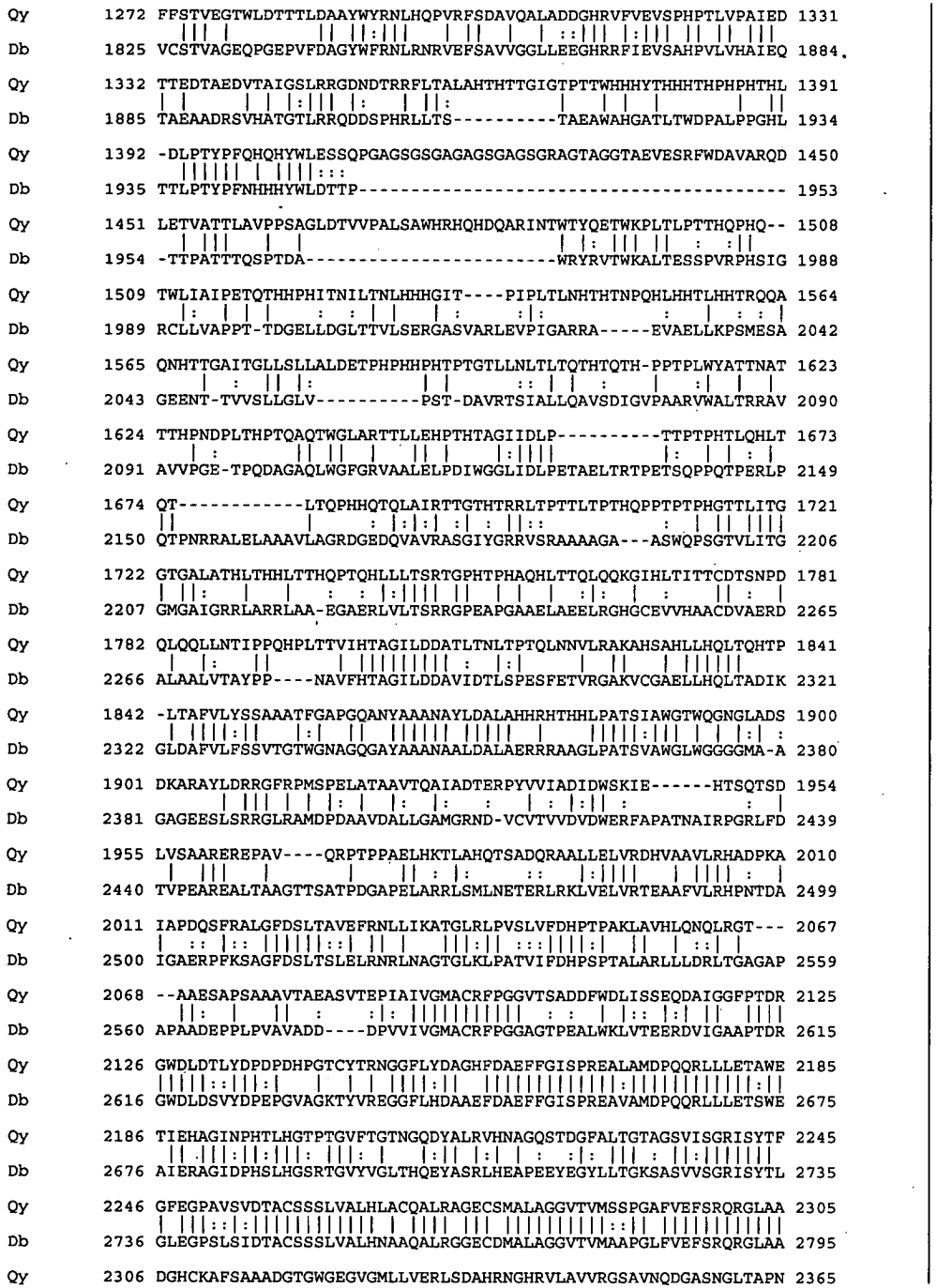
Qy        977 AEALRQYVAVRPDVSLADVGAGLACGRAVLEHRAVVLAADREELVQGLGALAAGEPDRRV  1036
              | | : |      | ||||   :| ||: . :||||||     ||    | ||| |:||  |
Db       1532 AGRLAEVVEA-GGVGLADVAVTMA-GRSRFGYRAVVLARGEAELAGRLRALAGGDPDAGV 1589

Qy       1037 TTGHA----PGGDRGGVVFVFPGQGGQWAGMGVRLLASSPVFARRMQACEEALAPWVDWS  1092
                ||        |  |||| |||||| || ||| || ||| |:  |  ||:  |:|
Db       1590 VTGAVVDPETGSGGGGVVLVFPGQGTQWVGMGAGLLGSSEVFAASMRECARALSVHVEWD 1649

Qy       1093 VVDILRRDAGDAVWERADVVQPVLFSVMVSLAALWRSYGIEPDAVLGHSQGEIAAAHVCG  1152
              :::::   :| |  || |||||| ::||||||  |:: |::  ||:|||||||||| | |
Db       1650 LLEVV---SGGAGLERVDVVQPVTWAVMVSLARYWQAMGVDVAAVVGHSQGEIAAATVAG 1706

Qy       1153 ALSLKDAAKTVALRSRALAA-VRGRGGMASVPLPAQEVEQLIGERWAGRLWVAAVNGPRS  1211
              ||||:|||  ||||:  :   : ||| ||:||||| |||  :  :| | : |||||| |
Db       1707 ALSLEDAAAVVALRAGLIGRYLAGRGAMAAVPLPAGEVEAGLA-KWPG-VEVAAVNGPAS 1764

Qy       1212 TAVSGDAEAVDEVLAYCAGTGVRARRIPVDYASHCPHVQPLREELLELLGDISPQPSGVP  1271
              | ||||  ||   :| |   ||:|| |||||||  ||: |: ||  :|  | |:   ||
Db       1765 TVVSGDRRAVAGYVAVCQAEGVQARLIPVDYASHSRHVEDLKGELERVLSGIRPRSPRVP 1824
```


Qy 1272 FFSTVEGTWLDTTTLDAAYWYRNLHQPVRFSDAVQALADDGHRVFVEVSPHPTLVPAIED 1331
Db 1825 VCSTVAGEQPGEPVFDAGYWFRNLRNRVEFSAVVGGLLEEGHRRFIEVSAHPVLVHAIEQ 1884
Qy 1332 TTEDTAEDVTAIGSLRRGDNDTRRFLTALAHTHTTGIGTPTTWHHHYTHHHTHPHPHTHL 1391
Db 1885 TAEAADRSVHATGTLRRQDDSPHRLLTS----------TAEAWAHGATLTWDPALPPGHL 1934
Qy 1392 -DLPTYPFQHQHYWLESSQPGAGSGSGAGAGSGAGSGRAGTAGGTAEVESRFWDAVARQD 1450
Db 1935 TTLPTYPFNHHHYWLDTTP---------- 1953
Qy 1451 LETVATTLAVPPSAGLDTVVPALSAWHRHQHDQARINTWTYQETWKPLTLPTTHQPHQ-- 1508
Db 1954 -TTPATTTQSPTDA----------WRYRVTWKALTESSPVRPHSIG 1988
Qy 1509 TWLIAIPETQTHHPHITNILTNLHHHGIT----PIPLTLNHTHTNPQHLHHTLHHTRQQA 1564
Db 1989 RCLLVAPPT-TDGELLDGLTTVLSERGASVARLEVPIGARRA-----EVAELLKPSMESA 2042
Qy 1565 QNHTTGAITGLLSLLALDETPHPHHPHTPTGTLLNLTLTQTHTQTH-PPTPLWYATTNAT 1623
Db 2043 GEENT-TVVSLLGLV----------PST-DAVRTSIALLQAVSDIGVPAARVWALTRRAV 2090
Qy 1624 TTHPNDPLTHPTQAQTWGLARTTLLEHPTHTAGIIDLP----------TTPTPHTLQHLT 1673
Db 2091 AVVPGE-TPQDAGAQLWGFGRVAALELPDIWGGLIDLPETAELTRTPETSQPPQTPERLP 2149
Qy 1674 QT----------LTQPHHQTQLAIRTTGTHTRRLTPTTLTPTHQPPTPTPHGTTLITG 1721
Db 2150 QTPNRRALELAAAVLAGRDGEDQVAVRASGIYGRRVSRAAAAGA---ASWQPSGTVLITG 2206
Qy 1722 GTGALATHLTHHLTTHQPTQHLLLTSRTGPHTPHAQHLTTQLQQKGIHLTITTCDTSNPD 1781
Db 2207 GMGAIGRRLARRLAA-EGAERLVLTSRRGPEAPGAAELAEELRGHGCEVVHAACDVAERD 2265
Qy 1782 QLQQLLNTIPPQHPLTTVIHTAGILDDATLTNLTPTQLNNVLRAKAHSAHLLHQLTQHTP 1841
Db 2266 ALAALVTAYPP----NAVFHTAGILDDAVIDTLSPESFETVRGAKVCGAELLHQLTADIK 2321
Qy 1842 -LTAFVLYSSAAATFGAPGQANYAAANAYLDALAHHRHTHHLPATSIAWGTWQGNGLADS 1900
Db 2322 GLDAFVLFSSVTGTWGNAGQGAYAAANAALDALAERRRAAGLPATSVAWGLWGGGGMA-A 2380
Qy 1901 DKARAYLDRRGFRPMSPELATAAVTQAIADTERPYVVIADIDWSKIE------HTSQTSD 1954
Db 2381 GAGEESLSRRGLRAMDPDAAVDALLGAMGRND-VCVTVVDVDWERFAPATNAIRPGRLFD 2439
Qy 1955 LVSAARAREPAV----QRPTPPAELHKTLAHQTSADQRAALLELVRDHVAAVLRHADPKA 2010
Db 2440 TVPEAREALTAAGTTSATPDGAPELARRLSMLNETERLRKLVELVRTEAAFVLRHPNTDA 2499
Qy 2011 IAPDQSFRALGFDSLTAVEFRNLLIKATGLRLPVSLVFDHPTPAKLAVHLQNQLRGT--- 2067
Db 2500 IGAERPFKSAGFDSLTSLELRNRLNAGTGLKLPATVIFDHPSPTALARLLLDRLTGAGAP 2559
Qy 2068 --AAESAPSAAAVTAEASVTEPIAIVGMACRFPGGVTSADDFWDLISSEQDAIGGFPTDR 2125
Db 2560 APAADEPPLPVAVADD----DPVVIVGMACRFPGGAGTPEALWKLVTEERDVIGAAPTDR 2615
Qy 2126 GWDLDTLYDPDPDHPGTCYTRNGGFLYDAGHFDAEFFGISPREALAMDPQQRLLLETAWE 2185
Db 2616 GWDLDSVYDPEPGVAGKTYVREGGFLHDAAEFDAEFFGISPREAVAMDPQQRLLLETSWE 2675
Qy 2186 TIEHAGINPHTLHGTPTGVFTGTNGQDYALRVHNAGQSTDGFALTGTAGSVISGRISYTF 2245
Db 2676 AIERAGIDPHSLHGSRTGVYVGLTHQEYASRLHEAPEEYEGYLLTGKSASVVSGRISYTL 2735
Qy 2246 GFEGPAVSVDTACSSSLVALHLACQALRAGECSMALAGGVTVMSSPGAFVEFSRQRGLAA 2305
Db 2736 GLEGPSLSIDTACSSSLVALHNAAQALRGGECDMALAGGVTVMAAPGLFVEFSRQRGLAA 2795
Qy 2306 DGHCKAFSAAADGTGWGEGVGMLLVERLSDAHRNGHRVLAVVRGSAVNQDGASNGLTAPN 2365
Db 2796 DGRCKAFADGADGTAWAEGAGVVLVERLSDARRLGHPVLAVVCGSAVNQDGASNGLTAPS 2855
Qy 2366 GPSQQRVIRQALANAGLSAGDVDAVEAHGTGTTLGDPIEAQALLATYGQDRAGEGPLWLG 2425
Db 2856 GPSQERVIRQALANARLTVADVDVVEAHGTGTRLGDPIEAQALLGTYGRDRDAECPVWLG 2915
Qy 2426 SVKSNVGHTQAAAGVAGVIKMVMALRHGLLPRTLHVDEPSPHVDWSAGAVQLLTETVPWP 2485
Db 2916 SLKSNIGHAQAAAGVAGVIKMVLAMRYGWLPRTLHVDEPSRHVDWSAGGVRLLTEAREWP 2975
Qy 2486 GGEGRLRRAGVSSFGVSGTNAHVILEEAP-------ADDVPGGPPAGEG---------- 2527
Db 2976 GVD-RPRRAAVSAFGVSGTNAHLIL-EAPEALEALEATDAPEAPEAPEAPDVTDVTEALE 3033
Qy 2528 --DAGSDDEA-AAGSP-------GVWPWLVSAKSQPALRAQAQALHAHLTDHPGLDLADV 2577
Db 3034 APDATEAEGAKAPGSPEEAQPAVGVVPVVVSGRSRVVVREAAGRL-AEVVEAGGVGLADV 3092
Qy 2578 GYTLAHARAVFDHRATLIAADRDTFLQALQALAAGEPHPAVIHSSAPGGTGTGEAAGKTA 2637
Db 3093 AVTMA-GRSRFGYRAVVLARGEAELAGRLRALAGGDPDAGVV-TGAVVDPETGSGGGGVV 3150
Qy 2638 FICSGQGTQRPGMAHGLYHTHPVFAAALNDICTHLDPHLDHPLLPLLTQNDNDNEDAAAL 2697
Db 3151 LVFPGQGTQWVGMGAGLLGSSEVFAASMRECARALSVHVEWDLLEVVS--------GGAG 3202
Qy 2698 LQQTRYAQPALFAFQVALHRLLTDGYHITPHYYAGHSLGEITAAHLAGILTLTDATTLIT 2757
Db 3203 LERVDVVQPVTWAVMVSLAR-YWQAMGVDVAAVVGHSQGEIAAATVAGALSLEDAAAVVA 3261
Qy 2758 QRATLM--QTMPPGTMTTLHTTPHHITHHLTAHENDLAIAAAINTPTSLVISGTPHTVQHI 2815
Db 3262 LRAGLIGRYLAGRGAMAAVPLPAGEVEAGL-AKWPGVQVAAVNGPASTVVSGDRRAVAGY 3320
Qy 2816 TTLCQQQGIKTKTLPTNHAFHSPHTNPILNQLHQHTQTLTYHPPHTPL---ITANTPPDQ 2872
Db 3321 VAVCQAEGVQARLIPVDYASHSRHVEDLKGELERVLSGIRPRSPRVPVCSTVAGEQPGEP 3380
Qy 2873 LLTPHYWTQQARNTVDYATTTQTLHQHGVTTYIELGPDNTLT---TLTHHNLPNPPTTTL 2929
Db 3381 VFDAGYWFRNLRNRVEFSAVVGGLLEQGHRRFIEVSAHPVLVHAIEQTAEAADRSVHATG 3440
Qy 2930 TLTHPHHHPQTHLLTNLAK-----TTTTWH----PHHYTHHDNQPHTHTHLDLPTYPFQH 2980
Db 3441 TLRRQDDSPH-RLLTSTAEAWAHGATLTWDPALPPGHLT-----------TLPTYPFNH 3487
Qy 2981 HHYWLESTQPGAGNVSAA--GLDPTEHPLLGATLELATDGGALLAGRLSLRSHPWLADHA 3038
Db 3488 HHYWAVTSPAGVGDAAAGRFGMTWEDHPFLRGGLPLADSGERVFAGRLAGSEHDWLTDHA 3547
Qy 3039 VGGTVLLSGATFLELALHAGTYVGCDRVDELTLHAPLVVPVDGGVSVQVGVAAADGEGRR 3098
Db 3548 VSGVTLLPGTAFVEFALHAGAATGCGRLEELSVEAPLVLPAAGGVRVQMRVSAADESGRR 3607
Qy 3099 LVSVYARGGSA---CGGGGASGGVWTCHASGVLVEAAAGGVVVDGLAGVWPPRGAVAVDV 3155
Db 3608 RVAIHSAPEAAVHSAAEGGDSAGVWTRHGEGTLVPDPEPTPPDADWARAWPPAGE-RVEP 3666
Qy 3156 DGVRDRLAGAGCVLGPVFSGLRAVWRDGGDLLAEVCLPEEAWGDAAGFGLHPALLDGVVQ 3215
Db 3667 AELYERFGALGYEYGEAFAGVRAVWRQPDALLAEVLLPDRASTGAGRFGVHPALLDAALQ 3726
Qy 3216 PLSVLLPGG-TGFGEGAGFGEGVRVPAVWGGVSLHRAGVTGVRVRVSAVGRGGGREAVSV 3274
Db 3727 P---WIAGGLLEVPEDA-----VLLPFAWQGVSLYATGAGALRVRLTKAGDG----AVSL 3774
Qy 3275 VVGDEAGVPVASVDRLELRPVDMGQLRA-VSVSAGRRGSLYAVQ----------WAE 3320
Db 3775 QAADTSGAAVLSLGALVMRPLARRKLDVLLGTDAGER-SLYRVEWQPRLLPAGPPRSWAV 3833
Qy 3321 VGP-------VPVCGQAWAWHEDVGESGGGPV---PGVVVLRCPDAGAGG---------- 3360

```
Db       3834 LGPDADRLAGTPGLGD---------QPDGGPTALYPEVRALR--KALAAGAPRPEAVVLP 3882

Qy       3361 --GGGGGGGGGVGEVVGGVLGVVQGWLGLERFAGSRLVVVTRGAVVAGPEDGPVDVVGAS 3418

Db       3883 VLSGAGATPESVRQTTERCLTALQDWLDAEELVDTPLIVLTRGAVAAVPGEEIGDLACAG 3942

Qy       3419 VWGLVRSAQAEHPDRFVLLDLDTDTGTDLDTGAGAGWGVDGGRVAAVVACGEPQLAVRGE 3478

Db       3943 VWGLVRSARSEHPGRFALVDTD--------------GHPDDRTAL-------PLALRAV 3980

Qy       3479 RLLAARLKRLESSGDVPAQRSGDTRA---RRSDVPAQRSGGVPARRSVDVSGREVLPWLS 3535

Db       3981 ---------LDGAGQL-SLRAGTARTPVLLRAGTPEEQRG--PA------------FDP 4015

Qy       3536 GGSVLVTGGTGVLGAAVARHLAGVCGVRDLLLVSRRGPDAPGAEGLRAELAALGAEVRIV 3595

Db       4016 AGTVLVTGATGTLGRLLARHLAAEHGVRHLLLLSRGGRAAEGADELAAELAGLEAEPCFA 4075

Qy       3596 ACDVGERREVVRLLEGVPAGCPLTGVVHAAGVLDDATIASLTPERLGTVFAAKVDAALLL 3655

Db       4076 ACDAADREALARVLAEVPADRPLTGVIHAAGVLDDGTLDALTPERIGTVMRPKADAALNL 4135

Qy       3656 DELTRGMELSAFVLFSSAAGILGSAGQGNYAAANAALDALAYRRRAAGLPGVSLAWGLWE 3715

Db       4136 HELTRTSPLSVFAVFSGAAGILGRPGQANYAAANTFLDALAQHRRAHGLPAVSLAWGLWG 4195

Qy       3716 EASGMTGHLAGTDHRRIIRSGLHPMSTPDALALFDAALA--LDRPVLLPADLRPA----- 3768

Db       4196 GATGMTGHLSGTDLRRMRRSGIAPMTHDQGLALFDRALAASAEDPLLVPMRLDLAALVRE 4255

Qy       3769 ------PPLPPLLQDLLPATRR-RTTRTTTTGGAD---NGAQLHARLAGQTHEQQHTTLL 3818

Db       4256 RAEHGPDAVPGPLLGLLPARAAVRQAAAPVRGGAPAPAGGEGTAERLAGLGEEARLRELV 4315

Qy       3819 ALVRSHIATVLGHTTPDTIPPDRAFRDLGFDSLTAVELRNRLSRTTGLRLPTTLAFDHPN 3878

Db       4316 RLVRAEVSGVLGYSGPDAVEPGRPFKDLGFDSLTAVELRNRLGAATGLRLPTALVFDRPT 4375

Qy       3879 PTTLTHHLHTQLQPQPD--NAVAPVLAELDKLESALSALDKTDSASE----RVTLRLKSL 3932

Db       4376 SQAVAEYLAAELAGPRDGGDTAAAAFEGLEALAAAVGALAEDDLRRDVLRRRLTELAAAL 4435

Qy       3933 MLRWNAPQHPTAESADDDEKFTSATEAEIFKFIDNDL 3969

Db       4436 TPQGRNPSAPAPAPSDLDERLDSANDDDLFAFIEEQL 4472


RESULT 14
AAY39301
ID   AAY39301 standard; Protein; 5588 AA.
XX
AC   AAY39301;
XX
DT   01-DEC-1999  (first entry)
XX
DE   SpnE a polyketide synthase.
XX
KW   Spinosyn biosynthetic enzyme; open reading frame; ORF; insecticidal
KW   microlides; arachnid; nematode; insect; polyketide; polyketide synthase;
KW   PKS; extender module; initiator module; acyl transferase domain; AT;
KW   acyl carrier protein; ACP; beta-ketosynthase domain; KS; KR;
KW   dehydratase domain; DH; enoyl reductase domain; ER; beta-ketoreductase;
KW   insecticide.
XX
OS   Saccharopolyspora spinosa.
XX
FH   Key             Location/Qualifiers
FT   Domain          1..424
FT                   /label= KS8
FT                   /note= "Beta-ketosynthase domain: part of extender
FT                        module 8"
FT   Domain          530..848
FT                   /label= AT8
FT                   /note= "Acyl transferase domain: part of extender
FT                        module 8"
FT   Domain          883..1070
FT                   /label= DH8
FT                   /note= "Dehydratase domain: part of extender
FT                        module 8"
FT   Domain          1369..1552
FT                   /label= KR8
FT                   /note= "Beta-ketoreductase domain: part of extender
FT                        module 8"
FT   Domain          1648..1726
FT                   /label= ACP8
FT                   /note= "Acyl carrier protein domain: part of extender
FT                        module 8"
FT   Domain          1749..2173
FT                   /label= KS9
FT                   /note= "Beta-ketosynthase domain: part of extender
FT                        module 9"
FT   Domain          2286..2614
FT                   /label= AT9
FT                   /note= "Acyl transferase domain: part of extender
FT                        module 9"
FT   Domain          2640..2800
FT                   /label= DH9
FT                   /note= "Dehydratase domain: part of extender
FT                        module 9"
FT   Domain          3157..3341
FT                   /label= KR9
FT                   /note= "Beta-ketoreductase domain: part of extender
FT                        module 9"
FT   Domain          3422..3500
FT                   /label= ACP9
FT                   /note= "Acyl carrier protein domain: part of extender
FT                        module 9"
FT   Domain          3534..3948
FT                   /label= KS10
FT                   /note= "Beta-ketosynthase domain: part of extender
FT                        module 10"
FT   Domain          4060..4390
FT                   /label= AT10
FT                   /note= "Acyl transferase domain: part of extender
FT                        module 10"
FT   Domain          4413..4597
FT                   /label= DH10
FT                   /note= "Dehydratase domain: part of extender
FT                        module 10"
FT   Domain          4900..5078
FT                   /label= KR10
FT                   /note= "Beta-ketoreductase domain: part of extender
FT                        module 10"
FT   Domain          5172..5253
FT                   /label= ACP10
FT                   /note= "Acyl carrier protein domain: part of extender
FT                        module 10"
FT   Domain          5302..5555
FT                   /label= TE10
FT                   /note= "Thiolester domain: part of extender
FT                        module 10"
XX
PN   WO9946387-A1.
XX
PD   16-SEP-1999.
XX
PF   16-FEB-1999;   99WO-US03212.
XX
PR   09-MAR-1998;   98US-0036987.
XX
PA   (DOWC ) DOW AGROSCIENCES LLC.
XX
PI   Baltz RH,  Broughton MC,  Crawford KP,  Madduri K,  Merlo DJ;
PI   Treadway PJ,  Turner JR,  Waldron C;
XX
DR   WPI; 1999-551414/46.
```

```
DR   N-PSDB; AAZ21501.
XX
PT   New spinosyn biosynthetic genes from Saccharopolyspora spinosa, useful
PT   for production of insecticidal spinosyn compounds -
XX
PS   Claim 1; Page 130-148; 190pp; English.
XX
CC   This is the amino acid sequence of the product of the spnE gene. The
CC   protein is involved in spinosyn biosynthesis. The SpnE gene is one of 23
CC   genes and open reading frames contained in an 80kb DNA sequence
CC   AAZ21501. Spinosyns are insecticidal microlides which are useful for the
CC   control of arachnids, nematodes and insects. Biosynthesis of spinosyns
CC   occurs via stepwise condensation and modification of carboxylic acid
CC   precursors generating a linear polyketide which is modified further. The
CC   DNA sequence contains a central region of approximately 55kb which has
CC   homology to the DNA encoding the polyketide synthases (PKS) of known
CC   macrolide producers. The spinosyn PKS DNA region consists of 5 ORFs with
CC   stop codons at the end of acyl carrier protein (ACP) domains. Together
CC   the PKS polypeptides (AAY39297-Y39301) of which this sequence is one,
CC   form a complex consisting of an initiator module, spnA, and several
CC   extender modules spnB-spnE. Each extender module adds a specific acetyl
CC   Co-A precursor to a growing polyketide chain, and modifies the beta-keto
CC   group in a specific manner. A module in a PKS polypeptide consists of
CC   several domains with specific functions. The initiator module has an
CC   acyl transferase (AT) domain, and an acyl carrier protein (ACP) domain.
CC   The extender modules have the same domains plus a beta-ketosynthase (KS)
CC   domain and optionally a beta-ketoreductase domain, a dehydratase (DH)
CC   domain, and an enoyl reductase (ER) domain. The last extender module
CC   terminates with a thiolester domain. The products of the genes present
CC   in the upstream region the PKS genes have been assigned names spnF-spnS
CC   AAY39302-Y39315 and are responsible for different modifications in
CC   spinosyn biosynthesis. There are also two ORFs ORFL15 and ORFL16 present
CC   immediately upstream of spnS, producing polypeptides AAY39316-Y39317,
CC   and two ORFs ORF1 and ORF2 present downstream of the PKS region
CC   producing polypeptides AAY39318-Y39319. The genes are useful to improve
CC   yields of spinosyns, and for creating new spinosyns e.g. by mutagenesis,
CC   or interruption of steps in spinosyn biosynthesis. The modified
CC   spinosyns may be a new insect control agent or serve as substrates for
CC   further chemical modification and the creation of new semi-synthetic
CC   spinosyns. The genes are also useful to isolate similar sequences from S.
CC   spinosa or other species by hybridization.
XX
SQ   Sequence   5588 AA;

  Query Match          35.0%;  Score 7281.5;  DB 20;  Length 5588;
  Best Local Similarity   36.1%;  Pred. No. 0;
  Matches 1821;  Conservative  490;  Mismatches 1349;  Indels 1389;  Gaps   86;

Qy         12 PAAGEVLGVADEADGGVVFVFPGQGPQWPGMGRELLDASDVFRESVRACEAAFAPYVDWS 71
              |||  | | ||   | | |||||||  | || ||| ::| |||  :  | || |||||||
Db        546 PAANVVQGAAD-CKGKVAFVFPGQGSHWQGMARELSESSPVFRRKLAECAAATAPYVDWS 604

Qy         72 VEQVLRDSPDAPGLDRVDVVQPTLFAVMISLAALWRSQGVEPCAVLGHSLGEIAAAHVSG 131
              :  |||  |||| ||| ||:|  |||:|:||| |||| |||| ||:||| ||||||||:|
Db        605 LLGVLRGDPDAPALDRDDVIQLALFAMMVSLAELWRSCGVEPAAVVGHSQGEIAAAHVAG 664

Qy        132 GLSLADAARVVTLWSQAQTTLAGTGALVSVAATPDELLPRIAPWTEDNPARLAVAAVNGP 191
              |||  ||  |::     | :  | | | ::::|     :: |||    |      | |||||||
Db        665 ALSLTDAVRIIAARCDAVSALTGKGGMLAIALPESAVVKRIAGLPE-----LTVAAVNGP 719

Qy        192 RSTVVSGAREAVADLVADLTAAQVRTRMIPVDVPAHSPLMYAIEERVVSGLLPITPRPSR 251
              ||||||   |:  |  :|||  |:|| : :|  :||| :  :: |::  |  :   |:
Db        720 GSTVVSGEPSALERLQTELTAENVQTRRVGIDYASHSPQIAQVQGRLLDRLGEVGSEPAE 779

Qy        252 IPFHSSVTGGRLDTRELDAAYWYRNMSSTVRFEPAARLLLQQGPKTFVEMSPHPVLTMGL 311
              | |:|:||| | ||  ||| |||:|:   |||:     :  || : |||:||||:|| |:
Db        780 IAFYSTVTGERTDTGRLDADYWYQNLRQPVRFQQTVARMADQGYRFFVEVSPHPLLTAGI 839

Qy        312 QELAPDLGDTTGTADT--VIMGTLRRGQGTLDHFLTSLA--QLRG-------------- 352
              ||        |   ||   |::|:||||:|    :|||||  |:||
Db        840 QE-------TLEAADAGGVVVGSLRRGEGGSRRWLTSLAECQVRGLPVNWEQVFLNTGAR 892

Qy        353 ------------------------------------------------------------ 352

Db        893 RVPLPTYPFQRQRYWLESAEYDAGDLGSVGLLSAEHPLLGAAVTLADAGGFLLTGKLSVK 952

Qy        353 ------------------------------------------------------------ 352

Db        953 TQPWLADHVVGGAILLPGTAFVEMLIRAADQVGCDLIEELSLTTPLVLPATGAVQVQIAV 1012

Qy        353 ------------------------------------------------------------ 352

Db       1013 GGPDEAGRRSVRVHSCRDDAVPQDSWTCHATGTLTSSDHQDAGQGPDGIWPPNDAVAVPL 1072

Qy        353 ----------------------------------------HGE------------------ 355
                                                      | |
Db       1073 DSFYARAAERGFDFGPAFQGLQAAWKRGDEIFAEVGLPTAHREDAGRFGIHPALLDAALQ 1132

Qy        356 ---------------------------------------------------TSATTVLS--- 363
                                                                  |:  |||
Db       1133 ALGAAEEDPDEGWLPFAWQGVSLKATGALSLRVHLVPAGANAVSVFTTDTTGQAVLSIDS 1192

Qy        364 ------------------------------------------------------------ 363

Db       1193 LVLRQISDKQLAAARAMEHESLFRVDWKRISPGAAKPVSWAVIGNDELARACGSALGTEL 1252

Qy        364 ------------------------------------------------------------ 363

Db       1253 HPDLTGLADPPPDVVVVPCGASRQDLDVASEARAATQRMLDLIQDWLAAARFAGSRLVVV 1312

Qy        364 ------------------------------------------------------------ 363

Db       1313 TCGAASTGPAEGVSDLVHAASWGLLRSAQSENPDRFVLVDVDGTAESWRALAAAVRSGEP 1372

Qy        364 ------------------------------------------------------------ 363

Db       1373 QLALRAGEVRVPRLARCVAAEDSRIPVPGADGTVLISGGTGLLGGLVARHLVAERGVRRL 1432

Qy        364 ------------------------------------------------------------ 363

Db       1433 VLAGRRGWSAPGVTDLVDELVGLGAAVEVASCDVGDRAQLDRLLTTISAEFPLRGVVHAA 1492

Qy        364 ------------------------------------------------------------ 363

Db       1493 GALADGVVESLTPEHVAKVFGPKAAGAWHLHELTLDLDLSFFVLFSSFSGVAGAAGQGNY 1552

Qy        364 ------------------------------------------------------------ 363

Db       1553 AAANAFLDGLAQHRRTAGLPAVSLAWGLWEQPSGMTGALDAAGRSRIARTNPPMSAPDGL 1612

Qy        364 ------------------------------------------------------------ 363

Db       1613 RLFEMAFRVPGESLLVPVHVDLNALRADAADGGVPALLRDLVPAPVRRSAVNESADVNGL 1672

Qy        364 -ARLTALSPTQQQSLLLDLVRAHTMAVLNDDGNERTASDAGPSASFAHLGFDSVMGVELR 422
               ||  |    |:: || ||| |  |||   |    | : |   :|  |||||: ||| |
Db       1673 VGRLRRLPDLDQETQLLGLVREHVSAVLGHSG----AVEVGADRAFRDLGFDSLSGVEFR 1728

Qy        423 NRLSKATGLRLPVTLIFDHTTPAAVAARLRTAALGHLDEDTAPVPDSPSGHGGTAAADDP 482
              |||    |:||| | :||: || |:  .|    :| ::  | | | : :       ||||
Db       1729 NRLGGVLGVRLPATAVFDYPTPRALVRFLLDKLIGGVEAPT-PAPAAVA----AVTADDP 1783

Qy        483 IAIIGMACRFPGGVRSPKDLWELAASGGDAIGPFPTDRGWPTEQRHAQDPTQPGTFYPQG 542
              : |:|| ||:|||| ||::|| | | | ||:  || ||||          || : || |
Db       1784 VVIVGMGCRYPGGVSSPEELWRLVAGGLDAVAEFPDDRGWDQAGLFDPDPDRLGTSYVCE 1843

Qy        543 GGFLHDAAHFDAGFFGISPREALAMDPQQRLLLETSWEAFERAGIDPLSVRGSRTGVFAG 602
              |||| ||| ||||||||||||||||||||||| :||  |||||||||:||||||||||
Db       1844 GGFLRDAAEFDAGFFGISPREALAMDPQQRLLLEVAWETVERAGIDPLSLRGSRTGVFAG 1903

Qy        603 ALSFDYGPRMDTASSEGAADVEGHILTGTTGSVLSGRIAYSFGLEGPAITVDTGCSASLV 662
              :  ||| |  | : ||    ||::  |: | | |||:||||| ||||:|||| ||:|||
Db       1904 LMHHDYGARFITRAPEG---FEGYLGNGSAGGVFSGRVAYSFGFEGPAVTVDTACSSSLV 1960

Qy        663 TLHLACQSLRSGECTLALAGGVSVMSTLGMFIEFSRQRGLSVDGRCKAYSAAADGTGWGE 722
              |||| |:|||||| |||||||:||:| |||:||||||||: |||||:::||||||||||
```



```
Qy       2616 PAVIHSSAPGGTGTGEAAGKTAFICSGQGTQRPGMAHGLYHTHPVFAAALNDICTHLDPH 2675
Db       4075 ----HEAAGVVTGPVGIGGRTGFVFAGQGGQWLGMGRRLYSEFPAFAGAFDEACAELDAN 4130

Qy       2676 LDHP--LLPLLTQNDNDNEDAAALLQQTRYAQPALFAFQVALHRLLTDGYHITPHYYAGH 2733
Db       4131 LGREVGVRDVVFGSDE------SLLDRTLWAQSGLFALQVGLWELL-GTWGVRPSVVLGH 4183

Qy       2734 SLGEITAAHLAGILTLTDATTLITQRATLMQTMPP-GTMTTLHTTPHHITHHLTAHENDL 2792
Db       4184 SVGELAAAFAAGVLSMAEAARLVAGRARLMQALPSGGAMLAVSATEARVGPLLDGVRDRV 4243

Qy       2793 AIAAINTPTSLVISGTPHTVQHITTLCQQQGIKTKTLPTNHAFHSPHTNPILNQLHQHTQ 2852
Db       4244 GVAAVNAPGSVVLSGDRDVLDGIAGRLDGQGIRSRWLRVSHAFHSHRMDPMLAEFAELAR 4303

Qy       2853 TLTYHPPHTPLITANT----PPDQLLTPHYWTQQARNTVDYATTTQTLHQHGVTTYIELG 2908
Db       4304 SVDYRSPRLPIVSTLTGNLDDVGVMATPEYWVRQVREPVRFADGVQALVDQGVDTIVELG 4363

Qy       2909 PDNTLTTLTHHNLPNPPTTT-LTLTHPHHHPQTHLLTNLAKTTTTWHPHHYTHHDNQPHT 2967
Db       4364 PDGALSSLVQECVAESGRATGIPLVRRDRDEVRTVLDALAQT----------------HT 4407

Qy       2968 H---------------THLDLPTYPFQHHHYWLESTQPGAGNVSAAGLDPTEHPLLGATL 3012
Db       4408 RGGAVDWGSFFAGTRATQVDLPTYAFQRQRYWLEPSD--SGDVTGVGLTGAEHPLLGAVV 4465

Qy       3013 ELATDGGALLAGRLSLRSHPWLADHAVGGTVLLSGATFLELALHAGTYVGCDRVDELTLH 3072
Db       4466 PVAGGDEVLLTGRLSVGTHPWLAEHRVLGEVVVPGTALLEMAWRAGSQVGCERVEELTLE 4525

Qy       3073 APLVVPVDGGVSVQVGVAAADGEGRRLVSVYARGGSACGGGASGGVWTCHASGVLVEAA 3132
Db       4526 APLVLPERGAAAVQLAVGAPDEAGRRSLQLYSRGAD-------EDGDWRRIASGLLAQAN 4578

Qy       3133 AGGVVVDGLAGVWPPRGAVAVDVDGVRDRLAGAGCVLGPVFSGLRAVWRDGGDLLAEVCL 3192
Db       4579 A---VPPADSTAWPPDGAGQVDLAEFYERLAERGLTYGPVFQGLRAAWRHGDDIFAELA- 4634

Qy       3193 PEEAWGDAAGFGLHPALLDGVVQPLSVLLPGGTGFGEGAGFGEGVRVPAVWGGVSLHRAG 3252
Db       4635 ---GSPDASGFGIHPALLDAALHAMAL----------GASPDSEARLPFSWRGAQLYRAE 4681

Qy       3253 VTGVRVRVSAVGRGGGREAVSVVVGDEAGVPVASVDRLELRPVDMGQLRAVSVSAGRRGS 3312
Db       4682 GAALRVRLSPLGSG----AVSLTLVDATGRRVAAVESLSTRPVSTDQIGA---GRGDQER 4734

Qy       3313 LYAVQW---AEVGPVPVCGQAWA------WHEDVGESG----------------GGPV 3345
Db       4735 LLHVEWVRSAESAGMSLTSCAVVGLGEPEWHAALKTTGVQVESHADLASLATEVAKRGSA 4794

Qy       3346 PGVVVLRCPDAGAGGGGGGGGGGVGEV-------VGGVLGVVQGWLGLERFAGSRLVVV 3398
Db       4795 PGAVIVPCPRPRA----------MQELPTAARRATQQAMAMLQQWLADDRFVSTRLILL 4843

Qy       3399 TRGAVVAGPEDGPVDVVGASVWGLVRSAQAEHPDRFVLLDLDTDTGTDLDTGAGAGWGVD 3458
Db       4844 THRAVSAVAGEDVLDLVHAPLWGLVRSAQAEHPDRFALIDMDDERASQT---------- 4892

Qy       3459 GGRVAAVVACGEPQLAVRGERLLAARLKRLESSGDVPAQRSGDTRARRSDVPAQRSGGVP 3518
Db       4893 --ALAEALTAGEAQLAVRSGVVLAPRLGQ---------------------------- 4919

Qy       3519 ARRSVDVSGREVLPWLSGGSVLVTGGTGVLGAAVARHLAGVCGVRDLLLVSRRGPDAPGA 3578
Db       4920 ----VKVSGGEAFRW-DEGTVLVTGGTGGLGALLARHLVSAHGVRHLLLASRRGLAAPGA 4974

Qy       3579 EGLRAELAALGAEVRIVACDVGERREVVRLLEGVPAGCPLTGVVHAAGVLDDATIASLTP 3638
Db       4975 DELVAELEQAGADVAVVACDSADRDSLARLVASVPAENPLRVVVHAAGVLDDGVLMSMSP 5034

Qy       3639 ERLGTVFAAKVDAALLLDELTRGMELSAFVLFSSAAGILGSAGQGNYAAANAALDALAYR 3698

                                                                          |||  |   |||||  | ||||  : |||||||||  ||: | ||| |||| || |||||:
Db       5035 ERLDAVLRPKVDAAWYLHELTRELGLSAFVLFSSVAGLFGGAGQSNYAAGNAFLDALAHC 5094

Qy       3699 RRAAGLPGVSLAWGLWEEASGMTGHLAGTDHRRIIRSGLHPMSTPDALALFDAALALDRP 3758
Db       5095 RQAQGLPALSLASGLWASIDGMAGDLAAADVERLSRAGIGPLSAPGGLALFDAAVGSDEP 5154

Qy       3759 VLLPADL----------RPAPPLPPLLQDLLPATRRRTTRTTTTGGADNGAQLHARLAGQ 3808
Db       5155 LLAPVRLDVEALRVQARSVQTRIPEMLHGMAMGPSRRTPFTSRV------EPLHERLAGL 5208

Qy       3809 THEQQHTTLLALVRSHIATVLGHTTPDTIPPDRAFRDLGFDSLTAVELRNRLSRTTGLRL 3868
Db       5209 SEGERRQQVLQRVRADIAVVLGHGRSSDVDIEKPLAELGFDSLTAIELRNRLATATGLRL 5268

Qy       3869 PTTLAFDHPNPTTLTHHLHTQL----QPQP-----DNAVAPVLAELDKLESALSALDKTD 3919
Db       5269 PATLAFDHGTAAALAQHVCAQLGTATAPAPRRTDDNDATEPVRSLFQQAYAAGRILDGMD 5328

Qy       3920 SASERVTLR 3928
Db       5329 LVKVAAQLR 5337
```

RESULT 15
AAB70969

```
ID   AAB70969 standard; Protein; 5588 AA.
XX
AC   AAB70969;
XX
DT   28-AUG-2001  (first entry)
XX
DE   S. spinosa protein fragment encoded by ORF22, SEQ ID 50.
XX
KW   Forosamine; trimethylrhamnose; polyketide synthase; biosynthesis;
KW   spinosyn; polyketide aglycone; transgenic plant; insect resistance;
KW   macrolide; insecticidal; polyketide synthase.
XX
OS   Saccharopolyspora spinosa.
XX
PN   DE19957268-A1.
XX
PD   08-MAR-2001.
XX
PF   29-NOV-1999;   99DE-1057268.
XX
PR   27-AUG-1999;   99DE-1040596.
XX
PA   (FARB ) BAYER AG.
XX
PI   Eberz G,  Moehrle V,  Froede R,  Velten R,  Salas JA;
XX
DR   WPI; 2001-267102/28.
DR   N-PSDB; AAF88339.
XX
PT   New nucleic acid encoding enzymes for spinosyn biosynthesis, useful for
PT   recombinant production of insecticidal spinosyns and their derivatives
PT   -
XX
PS   Claim 57; Page 313-335; 354pp; German.
XX
CC   This invention describes a novel method nucleic acid (I) and its encoded
CC   polypeptide (II) containing at least one region that encodes an enzymatic
CC   activity involved in biosynthesis of spinosyns. (I) are used (i) to
CC   identify, inactivate or modulate genes involved in the biosynthesis of
CC   (II); (ii) to generate a library of polyketide synthases; (iii) for
CC   adding forosamine or trimethylrhamnose to a spinosyn or polyketide
CC   aglycone; and (iv) for recombinant production of the corresponding
CC   enzymes, which are used for production of (II), their precursors or
CC   derivatives, including production of transgenic plants that express (II)
CC   and thus have increased resistance to insects. (I) are also useful as
CC   markers for sequencing of the Saccharopolyspora spinosa genome. (II) are
CC   macrolides with insecticidal, but not antibacterial, activity, and can
```

```
CC   also be used to raise specific antibodies, useful for identifying
CC   expression clones in a gene bank. Cells transformed with (I) may produce
CC   (II) at significantly increased levels or produce new derivatives of
CC   (II). This sequence represents an S. spinosa polyketide synthase.
XX
SQ   Sequence   5588 AA;

  Query Match            35.0%;  Score 7281.5;  DB 22;  Length 5588;
  Best Local Similarity  36.1%;  Pred. No. 0;
  Matches 1821;  Conservative  490;  Mismatches 1349;  Indels 1389;  Gaps   86;

Qy         12 PAAGEVLGVADEADGGVVFVFPGQGPQWPGMGRELLDASDVFRESVRACEAAFAPYVDWS 71
Db        546 PAANVVQGAAD-CKGKVAFVFPGQGSHWQGMARELSESSPVFRRKLAECAAATAPYVDWS 604

Qy         72 VEQVLRDSPDAPGLDRVDVVQPTLFAVMISLAALWRSQGVEPCAVLGHSLGEIAAAHVSG 131
Db        605 LLGVLRGDPDAPALDRDDVIQLALFAMMVSLAELWRSCGVEPAAVVGHSQGEIAAAHVAG 664

Qy        132 GLSLADAARVVTLWSQAQTTLAGTGALVSVAATPDELLPRIAPWTEDNPARLAVAAVNGP 191
Db        665 ALSLTDAVRIIAARCDAVSALTGKGGMLAIALPESAVVKRIAGLPE-----LTVAAVNGP 719

Qy        192 RSTVVSGAREAVADLVADLTAAQVRTRMIPVDVPAHSPLMYAIEERVVSGLLPITPRPSR 251
Db        720 GSTVVSGEPSALERLQTELTAENVQTRRVGIDYASHSPQIAQVQGRLLDRLGEVGSEPAE 779

Qy        252 IPFHSSVTGGRLDTRELDAAYWYRNMSSTVRFEPAARLLLQQGPKTFVEMSPHPVLTMGL 311
Db        780 IAFYSTVTGERTDTGRLDADYWYQNLRQPVRFQQTVARMADQGYRFFVEVSPHPLLTAGI 839

Qy        312 QELAPDLGDTTGTADT--VIMGTLRRGQGTLDHFLTSLA--QLRG--------------- 352
Db        840 QE-------TLEAADAGGVVVGSLRRGEGGSRRWLTSLAECQVRGLPVNWEQVFLNTGAR 892

Qy        353 ------------------------------------------------------------ 352
Db        893 RVPLPTYPFQRQRYWLESAEYDAGDLGSVGLLSAEHPLLGAAVTLADAGGFLLTGKLSVK 952

Qy        353 ------------------------------------------------------------ 352
Db        953 TQPWLADHVVGGAILLPGTAFVEMLIRAADQVGCDLIEELSLTTPLVLPATGAVQVQIAV 1012

Qy        353 ------------------------------------------------------------ 352
Db       1013 GGPDEAGRRSVRVHSCRDDAVPQDSWTCHATGTLTSSDHQDAGQGPDGIWPPNDAVAVPL 1072

Qy        353 ----------------------------------------HGE----------------- 355
Db       1073 DSFYARAAERGFDFGPAFQGLQAAWKRGDEIFAEVGLPTAHREDAGRFGIHPALLDAALQ 1132

Qy        356 ------------------------------------------------TSATTVLS--- 363
Db       1133 ALGAAEEDPDEGWLPFAWQGVSLKATGALSLRVHLVPAGANAVSVFTTDTTGQAVLSIDS 1192

Qy        364 ------------------------------------------------------------ 363
Db       1193 LVLRQISDKQLAAARAMEHESLFRVDWKRISPGAAKPVSWAVIGNDELARACGSALGTEL 1252

Qy        364 ------------------------------------------------------------ 363
Db       1253 HPDLTGLADPPPDVVVVPCGASRQDLDVASEARAATQRMLDLIQDWLAAARFAGSRLVVV 1312

Qy        364 ------------------------------------------------------------ 363
Db       1313 TCGAASTGPAEGVSDLVHAASWGLLRSAQSENPDRFVLVDVDGTAESWRALAAAVRSGEP 1372

Qy        364 ------------------------------------------------------------ 363
Db       1373 QLALRAGEVRVPRLARCVAAEDSRIPVPGADGTVLISGGTGLLGGLVARHLVAERGVRRL 1432

Qy        364 ------------------------------------------------------------ 363
Db       1433 VLAGRRGWSAPGVTDLVDELVGLGAAVEVASCDVGDRAQLDRLLTTISAEFPLRGVVHAA 1492

Qy        364 ------------------------------------------------------------ 363
Db       1493 GALADGVVESLTPEHVAKVFGPKAAGAWHLHELTLDLDLSFFVLFSSFSGVAGAAGQGNY 1552

Qy        364 ------------------------------------------------------------ 363
Db       1553 AAANAFLDGLAQHRRTAGLPAVSLAWGLWEQPSGMTGALDAAGRSRIARTNPPMSAPDGL 1612

Qy        364 ------------------------------------------------------------ 363
Db       1613 RLFEMAFRVPGESLLVPVHVDLNALRADAADGGVPALLRDLVPAPVRRSAVNESADVNGL 1672

Qy        364 -ARLTALSPTQQQSLLLDLVRAHTMAVLNDDGNERTASDAGPSASFAHLGFDSVMGVELR 422
Db       1673 VGRLRRLPDLDQETQLLGLVREHVSAVLGHSG----AVEVGADRAFRDLGFDSLSGVEFR 1728

Qy        423 NRLSKATGLRLPVTLIFDHTTPAAVAARLRTAALGHLDEDTAPVPDSPSGHGGTAAADDP 482
Db       1729 NRLGGVLGVRLPATAVFDYPTPRALVRFLLDKLIGGVEAPT-PAPAAVA----AVTADDP 1783

Qy        483 IAIIGMACRFPGGVRSPKDLWELAASGGDAIGPFPTDRGWPTEQRHAQDPTQPGTFYPQG 542
Db       1784 VVIVGMGCRYPGGVSSPEELWRLVAGGLDAVAEFPDDRGWDQAGLFDPDPDRLGTSYVCE 1843

Qy        543 GGFLHDAAHFDAGFFGISPREALAMDPQQRLLLETSWEAFERAGIDPLSVRGSRTGVFAG 602
Db       1844 GGFLRDAAEFDAGFFGISPREALAMDPQQRLLLEVAWETVERAGIDPLSLRGSRTGVFAG 1903

Qy        603 ALSFDYGPRMDTASSEGAADVEGHILTGTTGSVLSGRIAYSFGLEGPAITVDTGCSASLV 662
Db       1904 LMHHDYGARFITRAPEG---FEGYLGNGSAGGVFSGRVAYSFGFEGPAVTVDTACSSSLV 1960

Qy        663 TLHLACQSLRSGECTLALAGGVSVMSTLGMFIEFSRQRGLSVDGRCKAYSAAADGTGWGE 722
Db       1961 ALHLAGQALRSGECDLALAGGVTVMATPGMFVEFSRQRGLAADGRCKSFAAAADGTGWGE 2020

Qy        723 GVGMLLVERLSDAVRLGHRVLAVVRGSAVNQDGASNGLTAPNGPAQERVIRQALANAGLS 782
Db       2021 GAGLVLLERLSDARRNGHAVLAVVRGSAVNQDGASNGLTAPNGPSQQRVITQALASAGLS 2080

Qy        783 VADVDVVEGHGTGTTLGDPIEAQALLATYGQ-RAGDRPLWLGSLKSNIGHTMAAAGVGGV 841
Db       2081 VSDVDAVEAHGTGTRLGDPIEAQALIATYGQGRDSDRPLWLGSVKSNIGHTQAAAGVAGV 2140

Qy        842 IKMVMALREGVLPRTLHVDKPSPQVDWSAGAVRLLTEAVPWPGDAAGRLRRAGVSSFGIG 901
Db       2141 IKMVMAMRHGQLPATLHVDEPTSEVDWSAGDVQLLTENTPWPGNS--HPRRVGVSSFGIS 2198

Qy        902 GTNAHVILEEAPAAGGCVAGGGVLEGAPGLAISVAESVAAPVAVSAPVAESVPVPVPVPV 961
Db       2199 GTNAHVILEQA----------------------SKTPDETADKSGPDSEST-VDLPA-V 2233

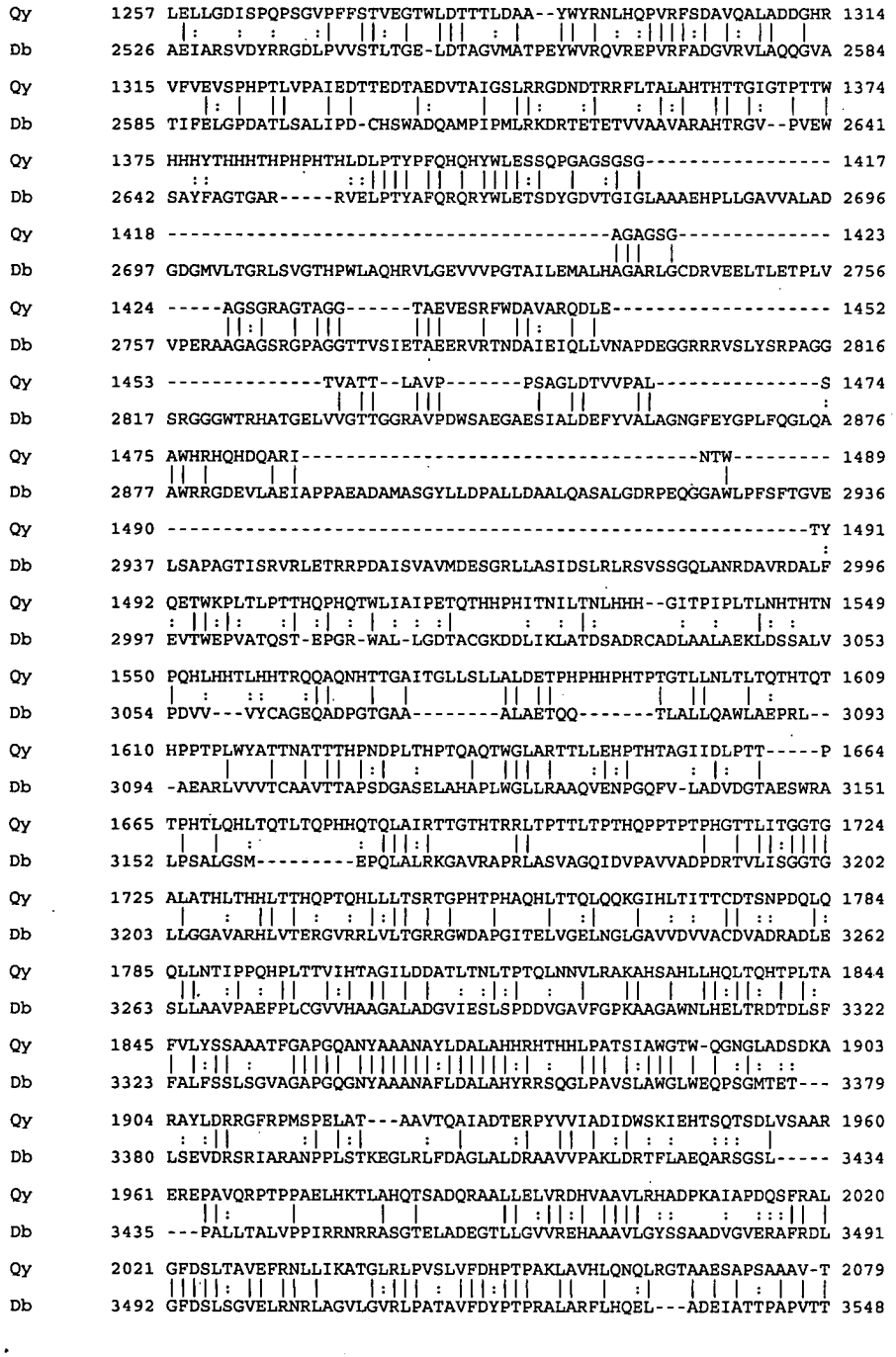
Qy        962 PVPVSARSEAGLRAQAEALRQYVAVRPDVSLADVGAGLACGRAVLEHRAVVLAADREELV 1021
Db       2234 PLIVSGRTPAALSAQASALLSYLGERGDISTLDAAFSLASSRAALEERAVVLGADRETLL 2293

Qy       1022 QGLGALAAGEPDRRVTTGHAPGGDRGGVVFVFPGQGGQWAGMGVRLLASSPVFARRMQAC 1081
Db       2294 SGLEALASGREASGVVSGSPVS---GGVGFVFAGQGGQWLGMGRGLYSVFPVFA---DAF 2347

Qy       1082 EEALAP-----WVDWSVVDILRRDAGDAVWERADVVQPVLFSVMVSLAALWRSYGIEPDA 1136
Db       2348 DEACAGLDAHLGQDVGVRDVVFGSDG-SLLDRTLWAQSGLFALQVGLLSLLGSWGVRPGV 2406

Qy       1137 VLGHSQGEIAAAHVCGALSLKDAAKTVALRSRALAAVRGRGGMASVPLPAQEVEQLIGER 1196
Db       2407 VLGHSVGEFAAAVAAGVLSLPDAARMVAGRARLMQALPSGGAMLAVAAGEEQLRPLLADR 2466

Qy       1197 WAGRLWVAAVNGPRSTAVSGDAEAVDEVLAYCAGTGVRARRIPVDYASHCPHVQPLREEL 1256
Db       2467 VDG-AGIAAVNAPESVVLSGDREVLDDIAGALDGQGIRWRRLRVSHAFHSYRMDPMLQEF 2525
```


Qy 1257 LELLGDISPQPSGVPFFSTVEGTWLDTTTLDAA--YWYRNLHQPVRFSDAVQALADDGHR 1314
Db 2526 AEIARSVDYRRGDLPVVSTLTGE-LDTAGVMATPEYWVRQVREPVRFADGVRVLAQQGVA 2584
Qy 1315 VFVEVSPHPTLVPAIEDTTEDTAEDVTAIGSLRRGDNDTRRFLTALAHTHTTGIGTPTTW 1374
Db 2585 TIFELGPDATLSALIPD-CHSWADQAMPIPMLRKDRTETETVVAAVARAHTRGV--PVEW 2641
Qy 1375 HHHYTHHHTHPHPHTHLDLPTYPFQHQHYWLESSQPGAGSGSG---------- 1417
Db 2642 SAYFAGTGAR-----RVELPTYAFQRQRYWLETSDYGDVTGIGLAAAEHPLLGAVVALAD 2696
Qy 1418 ----------AGAGSG---------- 1423
Db 2697 GDGMVLTGRLSVGTHPWLAQHRVLGEVVVPGTAILEMALHAGARLGCDRVEELTLETPLV 2756
Qy 1424 -----AGSGRAGTAGG------TAEVESRFWDAVARQDLE---------- 1452
Db 2757 VPERAAGAGSRGPAGGTTVSIETAEERVRTNDAIEIQLLVNAPDEGGRRRVSLYSRPAGG 2816
Qy 1453 ----------TVATT--LAVP-------PSAGLDTVVPAL----------S 1474
Db 2817 SRGGGWTRHATGELVVGTTGGRAVPDWSAEGAESIALDEFYVALAGNGFEYGPLFQGLQA 2876
Qy 1475 AWHRHQHDQARI----------NTW--------- 1489
Db 2877 AWRRGDEVLAEIAPPAEADAMASGYLLDPALLDAALQASALGDRPEQGGAWLPFSFTGVE 2936
Qy 1490 ----------TY 1491
Db 2937 LSAPAGTISRVRLETRRPDAISVAVMDESGRLLASIDSLRLRSVSSGQLANRDAVRDALF 2996
Qy 1492 QETWKPLTLPTTHQPHQTWLIAIPETQTHHPHITNILTNLHHH--GITPIPLTLNHTHTN 1549
Db 2997 EVTWEPVATQST-EPGR-WAL-LGDTACGKDDLIKLATDSADRCADLAALAEKLDSSALV 3053
Qy 1550 PQHLHHTLHHTRQQAQNHTTGAITGLLSLLALDETPHPHHPHTPTGTLLNLTLTQTHTQT 1609
Db 3054 PDVV---VYCAGEQADPGTGAA--------ALAETQQ-------TLALLQAWLAEPRL-- 3093
Qy 1610 HPPTPLWYATTNATTTHPNDPLTHPTQAQTWGLARTTLLEHPTHTAGIIDLPTT-----P 1664
Db 3094 -AEARLVVVTCAAVTTAPSDGASELAHAPLWGLLRAAQVENPGQFV-LADVDGTAESWRA 3151
Qy 1665 TPHTLQHLTQTLTQPHHQTQLAIRTTGTHTRRLTPTTLTPTHQPPTPTPHGTTLITGGTG 1724
Db 3152 LPSALGSM---------EPQLALRKGAVRAPRLASVAGQIDVPAVVADPDRTVLISGGTG 3202
Qy 1725 ALATHLTHHLTTHQPTQHLLLTSRTGPHTPHAQHLTTQLQQKGIHLTITTCDTSNPDQLQ 1784
Db 3203 LLGGAVARHLVTERGVRRLVLTGRRGWDAPGITELVGELNGLGAVVDVVACDVADRADLE 3262
Qy 1785 QLLNTIPPQHPLTTVIHTAGILDDATLTNLTPTQLNNVLRAKAHSAHLLHQLTQHTPLTA 1844
Db 3263 SLLAAVPAEPPLCGVVHAAGALADGVIESLSPDDVGAVFGPKAAGAWNLHELTRDTDLSF 3322
Qy 1845 FVLYSSAAATFGAPGQANYAAANAYLDALAHHRHTHHLPATSIAWGTW-QGNGLADSDKA 1903
Db 3323 FALFSSLSGVAGAPGQGNYAAANAFLDALAHYRRSQGLPAVSLAWGLWEQPSGMTET--- 3379
Qy 1904 RAYLDRRGFRPMSPELAT---AAVTQAIADTERPYVVIADIDWSKIEHTSQTSDLVSAAR 1960
Db 3380 LSEVDRSRIARANPPLSTKEGLRLFDAGLALDRAAVVPAKLDRTFLAEQARSGSL----- 3434
Qy 1961 EREPAVQRPTPPAELHKTLAHQTSADQRAALLELVRDHVAAVLRHADPKAIAPDQSFRAL 2020
Db 3435 ---PALLTALVPPIRRNRRASGTELADEGTLLGVVREHAAAVLGYSSAADVGVERAFRDL 3491
Qy 2021 GFDSLTAVEFRNLLIKATGLRLPVSLVFDHPTPAKLAVHLQNQLRGTAAESAPSAAAV-T 2079
Db 3492 GFDSLSGVELRNRLAGVLGVRLPATAVFDYPTPRALARFLHQEL---ADEIATTPAPVTT 3548
Qy 2080 AEASVTEP--IAIVGMACRFPGGVTSADDFWDLISSEQDAIGGFPTDRGWDLDTLYDPDP 2137
Db 3549 TRAPVAEDDLVAIVGMGCRFPGQVSSPEELWRLVAGGVDAVADFPADRGWDLAGLFDPDP 3608
Qy 2138 DHPGTCYTRNGGFLYDAGHFDAEFFGISPREALAMDPQQRLLLETAWETIEHAGINPHTL 2197
Db 3609 ERAGKTYVREGAFLTDADRFDAGFFGISPREALAMDPQQRLLLELSWEAIERAGIDPGSL 3668
Qy 2198 HGTPTGVFTGTNGQDYALR-VHNAGQSTDGFALTGTAGSVISGRISYTFGFEGPAVSVDT 2256
Db 3669 RGSRTGVFAGLMYHDYGARFASRAPEGFEGYLGNGSAGSVASGRIAYSFGFEGPAVTVDT 3728
Qy 2257 ACSSSLVALHLACQALRAGECSMALAGGVTVMSSPGAFVEFSRQRGLAADGHCKAFSAAA 2316
Db 3729 ACSSSLVALHLAGQSLRSGECDLALAGGVTVMSTPGTFVEFSRQRGLAPDGRCKSFAESA 3788
Qy 2317 DGTGWGEGVGMLLVERLSDAHRNGHRVLAVVRGSAVNQDGASNGLTAPNGPSQQRVIRQA 2376
Db 3789 DGTGWGEGAGLVLLERLSDARRNGHRVLAVVRGSAVNQDGASNGLTAPNGPSQQRVIQQA 3848
Qy 2377 LANAGLSAGDVDAVEAHGTGTTLGDPIEAQALLATYGQDRAGEGPLWLGSVKSNVGHTQA 2436
Db 3849 LASAGLSVSDVDAVEAHGTGTRLGDPIEAQALIATYGRDRDPGRPLWLGSVKSNIGHTQA 3908
Qy 2437 AAGVAGVIKMVMALRHGLLPRTLHVDEPSPHVDWSAGAVQLLTETVPWPGGEGRLRRAGV 2496
Db 3909 AAGVAGVIKMVMAMRHGQLPRTLHVDAPSSQVDWSAGRVQLLTENTPWP-DSGRPCRVGV 3967
Qy 2497 SSFGVSGTNAHVILEEAPADDVPGGPPAGEGD-AGSDDEAAAGSPGVWPWLVSAKSQPAL 2555
Db 3968 SSFGISGTNAHVILEQS----------TGQMDQAAEPDSSPVLDVPVVPWVVSGKTPEAL 4017
Qy 2556 RAQAQALHAHLTDHPGLDLADVGYTLAHARAVFDHRATLIAADRDTFLQALQALAAGEPH 2615
Db 4018 SAQAATLATYLDQNVDVSPLDVGISLAVTRSALDERAVVLGSDRDTLLSGLNALAAG--- 4074
Qy 2616 PAVIHSSAPGGTGTGEAAGKTAFICSGQGTQRPGMAHGLYHTHPVFAAALNDICTHLDPH 2675
Db 4075 ----HEAAGVVTGPVGIGGRTGFVFAGQGGQWLGMGRRLYSEFPAFAGAFDEACAELDAN 4130
Qy 2676 LDHP--LLPLLTQNDNDNEDAAALLQQTRYAQPALFAFQVALHRLLTDGYHITPHYYAGH 2733
Db 4131 LGREVGVRDVVFGSDE------SLLDRTLWAQSGLFALQVGLWELL-GTWGVRPSVVLGH 4183
Qy 2734 SLGEITAAHLAGILTLTDATTLITQRATLMQTMPP-GTMTTLHTTPHHITHHLTAHENDL 2792
Db 4184 SVGELAAAFAAGVLSMAEAARLVAGRARLMQALPSGGAMLAVSATEARVGPLLDGVRDRV 4243
Qy 2793 AIAAINTPTSLVISGTPHTVQHITTLCQQQGIKTKTLPTNHAFHSPHTNPILNQLHQHTQ 2852
Db 4244 GVAAVNAPGSVVLSGDRDVLDGIAGRLDGQGIRSRWLRVSHAFHSHRMDPMLAEFAELAR 4303
Qy 2853 TLTYHPPHTPLITANT----PPDQLLTPHYWTQQARNTVDYATTTQTLHQHGVTTYIELG 2908
Db 4304 SVDYRSPRLPIVSTLTGNLDDVGVMATPEYWVRQVREPVRFADGVQALVDQGVDTIVELG 4363
Qy 2909 PDNTLTTLTHHNLPNPPTTT-LTLTHPHHHPQTHLLTNLAKTTTTWHPHHYTHHDNQPHT 2967
Db 4364 PDGALSSLVQECVAESGRATGIPLVRRDRDEVRTVLDALAQT----------------HT 4407
Qy 2968 H----------THLDLPTYPFQHHHYWLESTQPGAGNVSAAGLDPTEHPLLGATL 3012
Db 4408 RGGAVDWGSFFAGTRATQVDLPTYAFQRQRYWLEPSD--SGDVTGVGLTGAEHPLLGAVV 4465
Qy 3013 ELATDGGALLAGRLSLRSHPWLADHAVGGTVLLSGATFLELALHAGTYVGCDRVDELTLH 3072
Db 4466 PVAGGDEVLLTGRLSVGTHPWLAEHRVLGEVVVPGTALLEMAWRAGSQVGCERVEELTLE 4525
Qy 3073 APLVVPVDGGVSVQVGVAAADGEGRRLVSVYARGGSACGGGGASGGVWTCHASGVLVEAA 3132
Db 4526 APLVLPERGAAAVQLAVGAPDEAGRRSLQLYSRGAD-------EDGDWRRIASGLLAQAN 4578
Qy 3133 AGGVVVDGLAGVWPPRGAVAVDVDGVRDRLAGAGCVLGPVFSGLRAVWRDGGDLLAEVCL 3192

```
Db       4579 A---VPPADSTAWPPDGAGQVDLAEFYERLAERGLTYGPVFQGLRAAWRHGDDIFAELA- 4634

Qy       3193 PEEAWGDAAGFGLHPALLDGVVQPLSVLLPGGTGFGEGAGFGEGVRVPAVWGGVSLHRAG 3252

Db       4635 ---GSPDASGFGIHPALLDAALHAMAL----------GASPDSEARLPFSWRGAQLYRAE 4681

Qy       3253 VTGVRVRVSAVGRGGGREAVSVVVGDEAGVPVASVDRLELRPVDMGQLRAVSVSAGRRGS 3312

Db       4682 GAALRVRLSPLGSG----AVSLTLVDATGRRVAAVESLSTRPVSTDQIGA---GRGDQER 4734

Qy       3313 LYAVQW---AEVGPVPVCGQAWA------WHEDVGESG----------------GGPV 3345

Db       4735 LLHVEWVRSAESAGMSLTSCAVVGLGEPEWHAALKTTGVQVESHADLASLATEVAKRGSA 4794

Qy       3346 PGVVVLRCPDAGAGGGGGGGGGGVGEV-------VGGVLGVVQGWLGLERFAGSRLVVV 3398

Db       4795 PGAVIVPCPRPRA----------MQELPTAARRATQQAMAMLQQWLADDRFVSTRLILL 4843

Qy       3399 TRGAVVAGPEDGPVDVVGASVWGLVRSAQAEHPDRFVLLDLDTDTGTDLDTGAGAGWGVD 3458

Db       4844 THRAVSAVAGEDVLDLVHAPLWGLVRSAQAEHPDRFALIDMDDERASQT---------- 4892

Qy       3459 GGRVAAVVACGEPQLAVRGERLLAARLKRLESSGDVPAQRSGDTRARRSDVPAQRSGGVP 3518

Db       4893 --ALAEALTAGEAQLAVRSGVVLAPRLGQ---------------------------- 4919

Qy       3519 ARRSVDVSGREVLPWLSGGSVLVTGGTGVLGAAVARHLAGVCGVRDLLLVSRRGPDAPGA 3578

Db       4920 ----VKVSGGEAFRW-DEGTVLVTGGTGGLGALLARHLVSAHGVRHLLLASRRGLAAPGA 4974

Qy       3579 EGLRAELAALGAEVRIVACDVGERREVVRLLEGVPAGCPLTGVVHAAGVLDDATIASLTP 3638

Db       4975 DELVAELEQAGADVAVVACDSADRDSLARLVASVPAENPLRVVVHAAGVLDDGVLMSMSP 5034

Qy       3639 ERLGTVFAAKVDAALLLDELTRGMELSAFVLFSSAAGILGSAGQGNYAAANAALDALAYR 3698

Db       5035 ERLDAVLRPKVDAAWYLHELTRELGLSAFVLFSSVAGLFGGAGQSNYAAGNAFLDALAHC 5094

Qy       3699 RRAAGLPGVSLAWGLWEEASGMTGHLAGTDHRRIIRSGLHPMSTPDALALFDAALALDRP 3758

Db       5095 RQAQGLPALSLASGLWASIDGMAGDLAAADVERLSRAGIGPLSAPGGLALFDAAVGSDEP 5154

Qy       3759 VLLPADL----------RPAPPLPPLLQDLLPATRRRTTRTTTTGGADNGAQLHARLAGQ 3808

Db       5155 LLAPVRLDVEALRVQARSVQTRIPEMLHGMAMGPSRRTPFTSRV------EPLHERLAGL 5208

Qy       3809 THEQQHTTLLALVRSHIATVLGHTTPDTIPPDRAFRDLGFDSLTAVELRNRLSRTTGLRL 3868

Db       5209 SEGERRQQVLQRVRADIAVVLGHGRSSDVDIEKPLAELGFDSLTAIELRNRLATATGLRL 5268

Qy       3869 PTTLAFDHPNPTTLTHHLHTQL----QPQP-----DNAVAPVLAELDKLESALSALDKTD 3919

Db       5269 PATLAFDHGTAAALAQHVCAQLGTATAPAPRRTDDNDATEPVRSLFQQAYAAGRILDGMD 5328

Qy       3920 SASERVTLR 3928

Db       5329 LVKVAAQLR 5337


Search completed: June 17, 2003, 13:02:26
Job time : 334.578 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 17, 2003, 13:08:02 ; Search time 123.76 Seconds
(without alignments)
3431.399 Million cell updates/sec

Title: US-09-914-286-3
Perfect score: 20829
Sequence: 1 VQRMDGGEEPRPAAGEVLGV.....FTSATEAEIFKFIDNDLGLS 3972

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 408643 seqs, 106915682 residues

Total number of hits satisfying chosen parameters: 408643

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_AA:*
1: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
7: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep:*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*
10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep:*
11: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
12: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep:*
13: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
14: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

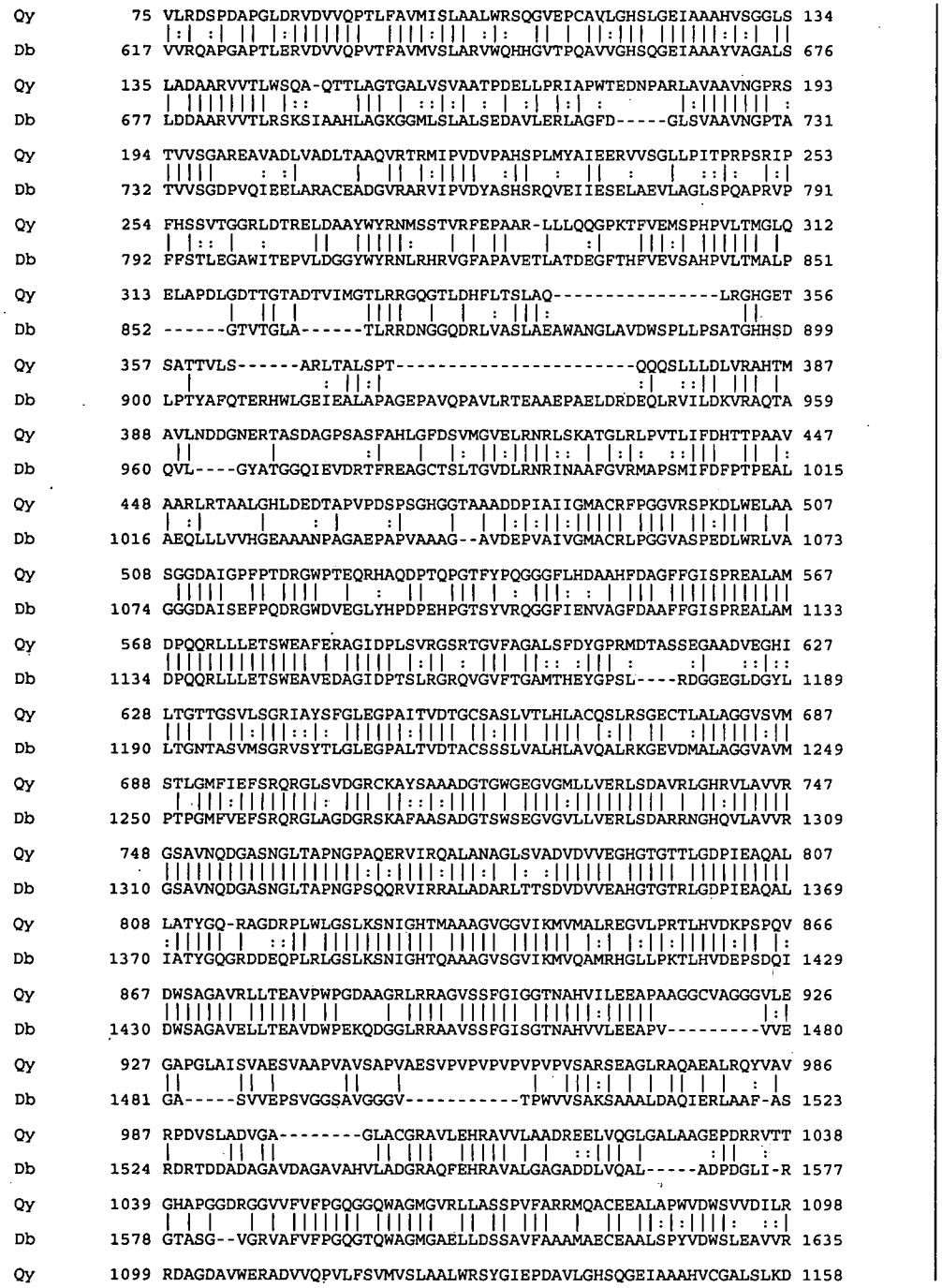
Result No.	Score	% Query Match	Length	DB	ID	Description
1	8151	39.1	4551	9	US-09-793-708-1	Sequence 1, Appli
2	8148	39.1	4613	9	US-09-860-846-31	Sequence 31, Appl
3	8148	39.1	4613	9	US-09-988-384B-31	Sequence 31, Appl
4	8148	39.1	4613	9	US-09-836-821-31	Sequence 31, Appl
5	8148	39.1	4613	10	US-09-861-289-31	Sequence 31, Appl
6	8148	39.1	11877	9	US-09-860-846-6	Sequence 6, Appli
7	8148	39.1	11877	9	US-09-836-821-6	Sequence 6, Appli
8	8148	39.1	11877	10	US-09-861-289-6	Sequence 6, Appli
9	8148	39.1	12199	9	US-09-988-384B-6	Sequence 6, Appli
10	7079	34.0	3739	9	US-09-793-708-2	Sequence 2, Appli
11	7041.5	33.8	3816	9	US-09-808-880-3	Sequence 3, Appli
12	7017	33.7	3739	9	US-09-860-846-33	Sequence 33, Appl
13	7017	33.7	3739	9	US-09-988-384B-33	Sequence 33, Appl
14	7017	33.7	3739	9	US-09-836-821-33	Sequence 33, Appl
15	7017	33.7	3739	10	US-09-861-289-33	Sequence 33, Appl
16	6912.5	33.2	4150	9	US-09-808-880-2	Sequence 2, Appli
17	6383	30.6	5215	9	US-09-860-846-2	Sequence 2, Appli
18	6383	30.6	5215	9	US-09-988-384B-2	Sequence 2, Appli
19	6383	30.6	5215	9	US-09-836-821-2	Sequence 2, Appli
20	6383	30.6	5215	10	US-09-861-289-2	Sequence 2, Appli
21	6108.5	29.3	3519	9	US-09-808-880-4	Sequence 4, Appli
22	5031	24.2	7257	9	US-10-014-717-5	Sequence 5, Appli
23	4695.5	22.5	3798	9	US-10-014-717-6	Sequence 6, Appli
24	3280.5	15.7	1562	9	US-09-860-846-35	Sequence 35, Appl
25	3280.5	15.7	1562	9	US-09-988-384B-35	Sequence 35, Appl
26	3280.5	15.7	1562	9	US-09-836-821-35	Sequence 35, Appl
27	3280.5	15.7	1562	9	US-09-793-708-3	Sequence 3, Appli
28	3280.5	15.7	1562	10	US-09-861-289-35	Sequence 35, Appl
29	2417.5	11.6	1346	9	US-09-793-708-4	Sequence 4, Appli
30	2412.5	11.6	1346	9	US-09-860-846-37	Sequence 37, Appl
31	2412.5	11.6	1346	9	US-09-988-384B-37	Sequence 37, Appl
32	2412.5	11.6	1346	9	US-09-836-821-37	Sequence 37, Appl
33	2412.5	11.6	1346	10	US-09-861-289-37	Sequence 37, Appl
34	2381	11.4	2439	9	US-10-014-717-7	Sequence 7, Appli
35	2142.5	10.3	1827	9	US-09-712-363-261	Sequence 261, App
36	2141.5	10.3	1832	9	US-10-014-717-4	Sequence 4, Appli
37	2029	9.7	1616	9	US-09-712-363-262	Sequence 262, App
38	1844.5	8.9	1421	9	US-10-014-717-2	Sequence 2, Appli
39	1732.5	8.3	1402	9	US-09-712-363-166	Sequence 166, App
40	1316	6.3	1610	9	US-09-738-626-6666	Sequence 6666, Ap
41	1293	6.2	458	10	US-09-892-870-4	Sequence 4, Appli
42	1280	6.1	2563	9	US-09-836-705-46	Sequence 46, Appl
43	1201	5.8	3032	9	US-09-836-705-44	Sequence 44, Appl
44	1194	5.7	1812	10	US-09-775-938A-38	Sequence 38, Appl
45	986.5	4.7	2509	9	US-10-237-271-1	Sequence 1, Appli

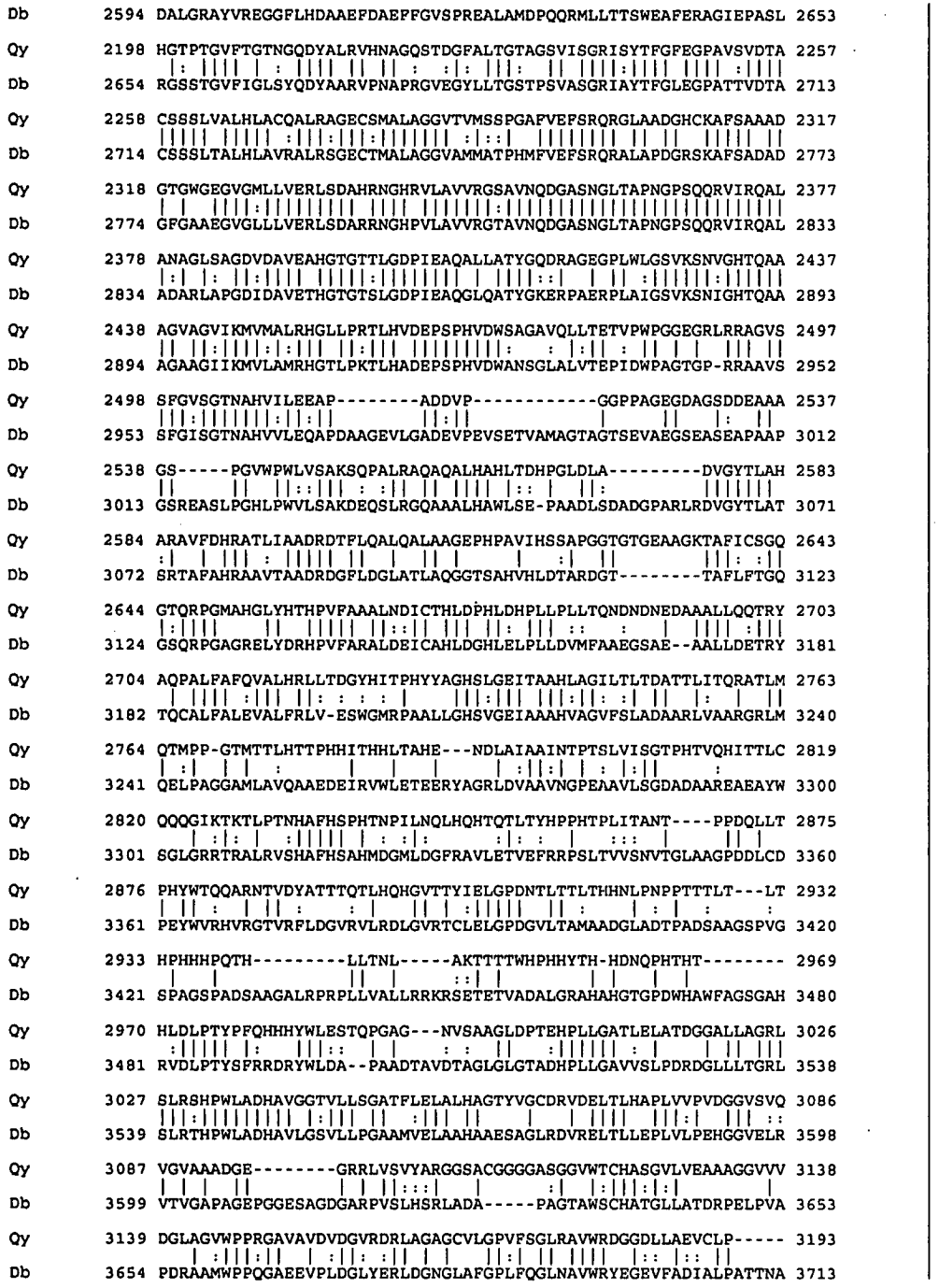
ALIGNMENTS

RESULT 1
US-09-793-708-1
; Sequence 1, Application US/09793708
; Publication No. US20030104597A1
; GENERAL INFORMATION:
; APPLICANT: ASHLEY, Gary
; APPLICANT: BETLACH, Melanie C.
; APPLICANT: BETLACH, Mary C.
; APPLICANT: McDANIEL, Robert
; APPLICANT: TANG, Li
; TITLE OF INVENTION: RECOMBINANT NARBONOLIDE POLYKETIDE SYNTHASE
; FILE REFERENCE: 300622002121
; CURRENT APPLICATION NUMBER: US/09/793,708
; CURRENT FILING DATE: 2001-02-22
; PRIOR APPLICATION NUMBER: US 09/657,440
; PRIOR FILING DATE: 2000-09-07
; PRIOR APPLICATION NUMBER: US 09/320,878
; PRIOR FILING DATE: 1999-05-27
; PRIOR APPLICATION NUMBER: US 09/141,908
; PRIOR FILING DATE: 1998-08-28
; PRIOR APPLICATION NUMBER: US 09/073,538
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: US 08/846,247
; PRIOR FILING DATE: 1997-04-30
; PRIOR APPLICATION NUMBER: US 60/134,990
; PRIOR FILING DATE: 1999-05-20
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 4551
; TYPE: PRT
; ORGANISM: Streptomyces venezuelae
US-09-793-708-1

Query Match 39.1%; Score 8151; DB 9; Length 4551;
Best Local Similarity 44.8%; Pred. No. 0;
Matches 1910; Conservative 462; Mismatches 1307; Indels 586; Gaps 94;

```
Qy         15 GEVLGVADEADGGVVFVFPGQGPQWPGMGRELLDASDVFRESVRACEAAFAPYVDWSVEQ 74
              | | |||    | | ||||||| || ||| ||||:| ||  ::  |||| :|||||:|
Db        558 GLVRGVASGV-GRVAFVFPGQGTQWAGMGAELLDSSAVFAAAMAECEAALSPYVDWSLEA 616
```





RESULT 2
US-09-860-846-31
; Sequence 31, Application US/09860846
; Patent No. US20020164742A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.

```
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/860,846
;  CURRENT FILING DATE:  2001-05-18
;  PRIOR APPLICATION NUMBER: 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 31
;   LENGTH: 4613
;   TYPE: PRT
;   ORGANISM: Streptomyces venezuelae
US-09-860-846-31

  Query Match             39.1%;  Score 8148;  DB 9;  Length 4613;
  Best Local Similarity   44.8%;  Pred. No. 0;
  Matches 1909;  Conservative  463;  Mismatches 1307;  Indels  586;  Gaps   94;

Qy         15 GEVLGVADEADGGVVFVFPGQGPQWPGMGRELLDASDVFRESVRACEAAFAPYVDWSVEQ  74
Db        620 GLVRGVASGV-GRVAFVFPGQGTQWAGMGAELLDSSAVFAAAMAECEAALSPYVDWSLEA  678

Qy         75 VLRDSPDAPGLDRVDVVQPTLFAVMISLAALWRSQGVEPCAVLGHSLGEIAAAHVSGGLS  134
Db        679 VVRQAPGAPTLERVDVVQPVTFAVMVSLARVWQHHGVTPQAVVGHSQGEIAAAYVAGALS  738

Qy        135 LADAARVVTLWSQA-QTTLAGTGALVSVAATPDELLPRIAPWTEDNPARLAVAAVNGPRS  193
Db        739 LDDAARVVTLRSKSIAAHLAGKGGMLSLALSEDAVLERLAGFD-----GLSVAAVNGPTA  793

Qy        194 TVVSGAREAVADLVADLTAAQVRTRMIPVDVPAHSPLMYAIEERVVSGLLPITPRPSRIP  253
Db        794 TVVSGDPVQIEELARACEADGVRARVIPVDYASHSRQVEIIESELAEVLAGLSPQAPRVP  853

Qy        254 PHSSVTGGRLDTRELDAAYWYRNMSSTVRFEPAAR-LLLQQGPKTFVEMSPHPVLTMGLQ  312
Db        854 FFSTLEGAWITEPVLDGGYWYRNLRHRVGFAPAVETLATDEGFTHFVEVSAHPVLTMALP  913

Qy        313 ELAPDLGDTTGTADTVIMGTLRRGQGTLDHFLTSLAQ----------------LRGHGET  356
Db        914 ------GTVTGLA------TLRRDNGGQDRLVASLAEAWANGLAVDWSPLLPSATGHHSD  961

Qy        357 SATTVLS------ARLTALSPT----------------------QQQSLLLDLVRAHTM  387
Db        962 LPTYAFQTERHWLGEIEALAPAGEPAVQPAVLRTEAAEPAELDRDEQLRVILDKVRAQTA 1021

Qy        388 AVLNDDGNERTASDAGPSASFAHLGFDSVMGVELRNRLSKATGLRLPVTLIFDHTTPAAV  447
Db       1022 QVL----GYATGGQIEVDRTFREAGCTSLTGVDLRNRINAAFGVRMAPSMIFDFPTPEAL 1077

Qy        448 AARLRTAALGHLDEDTAPVPDSPSGHGGTAAADDPIAIIGMACRFPGGVRSPKDLWELAA  507
Db       1078 AEQLLLVVHGEAAANPAGAEPAPVAAAG--AVDEPVAIVGMACRLPGGVASPEDLWRLVA 1135

Qy        508 SGGDAIGPFPTDRGWPTEQRHAQDPTQPGTFYPQGGGFLHDAAHFDAGFFGISPREALAM  567
Db       1136 GGGDAISEFPQDRGWDVEGLYHPDPEHPGTSYVRQGGFIENVAGFDAAFFGISPREALAM 1195

Qy        568 DPQQRLLLETSWEAFERAGIDPLSVRGSRTGVFAGALSFDYGPRMDTASSEGAADVEGHI  627
Db       1196 DPQQRLLLETSWEAVEDAGIDPTSLRGRQVGVFTGAMTHEYGPSL----RDGGEGLDGYL 1251

Qy        628 LTGTTGSVLSGRIAYSFGLEGPAITVDTGCSASLVTLHLACQSLRSGECTLALAGGVSVM  687
Db       1252 LTGNTASVMSGRVSYTLGLEGPALTVDTACSSSLVALHLAVQALRKGEVDMALAGGVAVM 1311

Qy        688 STLGMFIEFSRQRGLSVDGRCKAYSAAADGTGWGEGVGMLLVERLSDAVRLGHRVLAVVR  747
Db       1312 PTPGMFVEFSRQRGLAGDGRSKAFAASADGTSWSEGVGVLLVERLSDARRNGHQVLAVVR 1371

Qy        748 GSAVNQDGASNGLTAPNGPAQERVIRQALANAGLSVADVDVVEGHGTGTTLGDPIEAQAL  807
Db       1372 GSALNQDGASNGLTAPNGPSQQRVIRRALADARLTTSDVDVVEAHGTGTRLGDPIEAQAL 1431

Qy        808 LATYGQ-RAGDRPLWLGSLKSNIGHTMAAAGVGGVIKMVMALREGVLPRTLHVDKPSPQV  866
Db       1432 IATYGQGRDDEQPLRLGSLKSNIGHTQAAAGVSGVIKMVQAMRHGLLPKTLHVDEPSDQI 1491

Qy        867 DWSAGAVRLLTEAVPWPGDAAGRLRRAGVSSFGIGGTNAHVILEEAPAAGGCVAGGGVLE  926
Db       1492 DWSAGAVELLTEAVDWPEKQDGGLRRAAVSSFGISGTNAHVVLEEAPV---------VVE 1542

Qy        927 GAPGLAISVAESVAAPVAVSAPVAESVPVPVPVPVPVPVSARSEAGLRAQAEALRQYVAV  986
Db       1543 GA-----SVVEPSVGGSAVGGGV----------TPWVVSAKSAAALDAQIERLAAF-AS 1585

Qy        987 RPDVSLADVGA--------GLACGRAVLEHRAVVLAADREELVQGLGALAAGEPDRRVTT 1038
Db       1586 RDRTDDADAGAVDAGAVAHVLADGRAQFEHRAVALGAGADDLVQAL-----ADPDGLI-R 1639

Qy       1039 GHAPGGDRGGVVFVFPGQGGQWAGMGVRLLASSPVFARRMQACEEALAPWVDWSVVDILR  1098
Db       1640 GTASG--VGRVAFVFPGQGTQWAGMGAELLDSSAVFAAAMAECEAALSPYVDWSLEAVVR 1697

Qy       1099 RDAGDAVWERADVVQPVLFSVMVSLAALWRSYGIEPDAVLGHSQGEIAAAHVCGALSLKD  1158
Db       1698 QAPGAPTLERVDVVQPVTFAVMVSLARVWQHHGVTPQAVVGHSQGEIAAAYVAGALPLDD 1757

Qy       1159 AAKTVALRSRALAA-VRGRGGMASVPLPAQEVEQLIGERWAGRLWVAAVNGPRSTAVSGD  1217
Db       1758 AARVVTLRSKSIAAHLAGKGGMLSLALNEDAVLERLSD--FDGLSVAAVNGPTATVVSGD 1815

Qy       1218 AEAVDEVLAYCAGTGVRARRIPVDYASHCPHVQPLREELLELLGDISPQPSGVPFFSTVE  1277
Db       1816 PVQIEELAQACKADGFRARIIPVDYASHSRQVEIIESELAQVLAGLSPQAPRVPFFSTLE 1875

Qy       1278 GTWLDTTTLDAAYWYRNLHQPVRFSDAVQALA-DDGHRVFVEVSPHPTLVPAIEDTTEDT  1336
Db       1876 GTWITEPVLDGTYWYRNLRHRVGFAPAIETLAVDEGFTHFVEVSAHPVLTMTLPET---- 1931

Qy       1337 AEDVTAIGSLRRGDNDTRRFLTALAHTHTTGIGTPTTWHHHYTHHHTHPHPHTHLDLPTY  1396
Db       1932 ---VTGLGTLRREQGGQERLVTSLAEAWVNGL--PVAWTSLLPATASRP------GLPTY 1980

Qy       1397 PFQHQHYWLESSQPGAGSGSGAGAGSGAGSGRAGTAGGTAEVESRFWDAVARQDLETVAT  1456
Db       1981 AFQAERYWLENTPAALATG---------------------------------------- 1999

Qy       1457 TLAVPPSAGLDTVVPALSAWHRHQHDQARINTWTYQETWKPLTLPTTHQPHQT-----WL  1511
Db       2000 ----------------------------DDWRYRIDWK--RLPAAEGSERTGLSGRWL 2027

Qy       1512 IAIPETQTHHPHITNILTNLHHHGITPIPLTLNHTHTNPQHLHHTLHHTRQQAQNHTTG-  1570
Db       2028 AVTPE--DHSAQAAAVLTALVDAGAKVEVLTAG-ADDDREALAARLTAL-------TTGD 2077

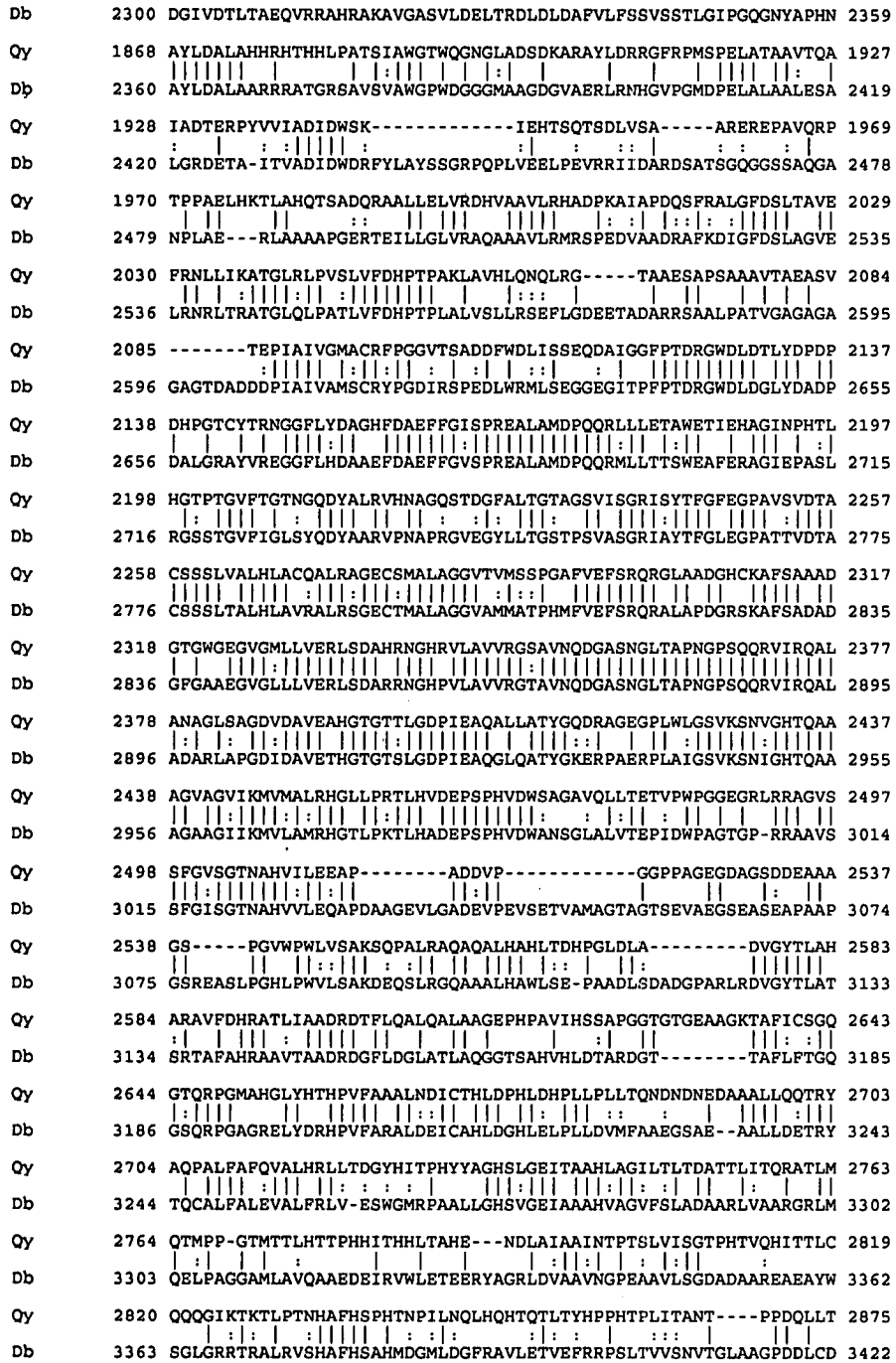
Qy       1571 AITGLLSLLALDETPHPHHPHTPTGTLLNLTLTQTHTQTHPPTPLWYATTNATTTHPNDP  1630
Db       2078 GFTGVVSLL--------------DGLVPQVAWVQALGDAGIKAPLWSVTQGAVSVGRLDT 2123

Qy       1631 LTHPTQAQTWGLARTTLLEHPTHTAGIIDLPTTPTPHTLQHLTQTLTQPHHQTQLAIRTT  1690
Db       2124 PADPDRAMLWGLGRVVALEHPERWAGLVDLPAQPDAAALAHLVTALSGATGEDQIAIRTT 2183

Qy       1691 GTHTRRLTPTTLTPTH-QPPTP--TPHGTTLITGGTGALATHLTHHLTTHQPTQHLLLTS  1747
Db       2184 GLHARRL---ARAPLHGRRPTRDWQPHGTVLITGGTGALGSHAARWM-AHHGAEHLLLVS 2239

Qy       1748 RTGPHTPHAQHLTTQLQQKGIHLTITTCDTSNPDQLQQLLNTIPPQHPLTTVIHTAGILD  1807
Db       2240 RSGEQAPGATQLTAELTASGARVTIAACDVADPHAMRTLLDAIPAETPLTAVVHTAGALD 2299

Qy       1808 DATLTNLTPTQLNNVLRAKAHSAHLLHQLTQHTPLTAFVLYSSAAATFGAPGQANYAAAN  1867
```


Db 2300 DGIVDTLTAEQVRRAHRAKAVGASVLDELTRDLDLDAFVLFSSVSSTLGIPGQGNYAPHN 2359
Qy 1868 AYLDALAHHRHTHHLPATSIAWGTWQGNGLADSDKARAYLDRRGFRPMSPELATAAVTQA 1927
Db 2360 AYLDALAARRRATGRSAVSVAWGPWDGGGMAAGDGVAERLRNHGVPGMDPELALAALESA 2419
Qy 1928 IADTERPYVVIADIDWSK------------IEHTSQTSDLVSA-----ARELREPAVQRP 1969
Db 2420 LGRDETA-ITVADIDWDRFYLAYSSGRPQPLVEELPEVRRIIDARDSATSGQGGSSAQGA 2478
Qy 1970 TPPAELHKTLAHQTSADQRAALLELVRDHVAAVLRHADPKAIAPDQSFRALGFDSLTAVE 2029
Db 2479 NPLAE---RLAAAAPGERTEILLGLVRAQAAAVLRMRSPEDVAADRAFKDIGFDSLAGVE 2535
Qy 2030 FRNLLIKATGLRLPVSLVFDHPTPAKLAVHLQNQLRG-----TAAESAPSAAAVTAEASV 2084
Db 2536 LRNRLTRATGLQLPATLVFDHPTPLALVSLLRSEFLGDEETADARRSAALPATVGAGAGA 2595
Qy 2085 -------TEPIAIVGMACRFPGGVTSADDFWDLISSEQDAIGGFPTDRGWDLDTLYDPDP 2137
Db 2596 GAGTDADDDPIAIVAMSCRYPGDIRSPEDLWRMLSEGGEGITPFPTDRGWDLDGLYDADP 2655
Qy 2138 DHPGTCYTRNGGFLYDAGHFDAEFFGISPREALAMDPQQRLLLETAWETIEHAGINPHTL 2197
Db 2656 DALGRAYVREGGFLHDAAEFDAEFFGVSPREALAMDPQQRMLLTTSWEAFERAGIEPASL 2715
Qy 2198 HGTPTGVFTGTNGQDYALRVHNAGQSTDGFALTGTAGSVISGRISYTFGFEGPAVSVDTA 2257
Db 2716 RGSSTGVFIGLSYQDYAARVPNAPRGVEGYLLTGSTPSVASGRIAYTFGLEGPATTVDTA 2775
Qy 2258 CSSSLVALHLACQALRAGECSMALAGGVTVMSSPGAFVEFSRQRGLAADGHCKAFSAAAD 2317
Db 2776 CSSSLTALHLAVRALRSGECTMALAGGVAMMATPHMFVEFSRQRALAPDGRSKAFSADAD 2835
Qy 2318 GTGWGEGVGMLLVERLSDAHRNGHRVLAVVRGSAVNQDGASNGLTAPNGPSQQRVIRQAL 2377
Db 2836 GFGAAEGVGLLLVERLSDARRNGHPVLAVVRGTAVNQDGASNGLTAPNGPSQQRVIRQAL 2895
Qy 2378 ANAGLSAGDVDAVEAHGTGTTLGDPIEAQALLATYGQDRAGEGPLWLGSVKSNVGHTQAA 2437
Db 2896 ADARLAPGDIDAVETHGTGTSLGDPIEAQGLQATYGKERPAERPLAIGSVKSNIGHTQAA 2955
Qy 2438 AGVAGVIKMVMALRHGLLPRTLHVDEPSPHVDWSAGAVQLLTETVPWPGGEGRLRRAGVS 2497
Db 2956 AGAAGIIKMVLAMRHGTLPKTLHADEPSPHVDWANSGLALVTEPIDWPAGTGP-RRAAVS 3014
Qy 2498 SFGVSGTNAHVILEEAP--------ADDVP------------GGPPAGEGDAGSDDEAAA 2537
Db 3015 SFGISGTNAHVVLEQAPDAAGEVLGADEVPEVSETVAMAGTAGTSEVAEGSEASEAPAAP 3074
Qy 2538 GS-----PGVWPWLVSAKSQPALRAQAQALHAHLTDHPGLDLA---------DVGYTLAH 2583
Db 3075 GSREASLPGHLPWVLSAKDEQSLRGQAAALHAWLSE-PAADLSDADGPARLRDVGYTLAT 3133
Qy 2584 ARAVFDHRATLIAADRDTFLQALQALAAGEPHPAVIHSSAPGGTGTGEAAGKTAFICSGQ 2643
Db 3134 SRTAFAHRAAVTAADRDGFLDGLATLAQGGTSAHVHLDTARDGT--------TAFLFTGQ 3185
Qy 2644 GTQRPGMAHGLYHTHPVFAAALNDICTHLDPHLDHPLLPLLTQNDNDNEDAAALLQQTRY 2703
Db 3186 GSQRPGAGRELYDRHPVFARALDEICAHLDGHLELPLLDVMFAAEGSAE--AALLDETRY 3243
Qy 2704 AQPALPAFQVALHRLLTDGYHITPHYYAGHSLGEITAAHLAGILTLTDATTLITQRATLM 2763
Db 3244 TQCALPALEVALFRLV-ESWGMRPAALLGHSVGEIAAAHVAGVFSLADAARLVAARGRLM 3302
Qy 2764 QTMPP-GTMTTLHTTPHHITHHLTAHE---NDLAIAAINTPTSLVISGTPHTVQHITTLC 2819
Db 3303 QELPAGGAMLAVQAAEDEIRVWLETEERYAGRLDVAAVNGPEAAVLSGDADAAREAEAYW 3362
Qy 2820 QQQGIKTKTLPTNHAFHSPHTNPILNQLHQHTQTLTYHPPHTPLITANT----PPDQLLT 2875
Db 3363 SGLGRRTRALRVSHAFHSAHMDGMLDGFRAVLETVEFRRPSLTVVSNVTGLAAGPDDLCD 3422
Qy 2876 PHYWTQQARNTVDYATTTQTLHQHGVTTYIELGPDNTLTTLTHHNLPNPPTTTLT---LT 2932
Db 3423 PEYWVRHVRGTVRFLDGVRVLRDLGVRTCLELGPDGVLTAMAADGLADTPADSAAGSPVG 3482
Qy 2933 HPHHHPQTH---------LLTNL-----AKTTTTWHPHHYTH-HDNQPHTHT-------- 2969
Db 3483 SPAGSPADSAAGALRPRPLLVALLRRKRSETETVADALGRAHAHGTGPDWHAWFAGSGAH 3542
Qy 2970 HLDLPTYPFQHHHYWLESTQPGAG---NVSAAGLDPTEHPLLGATLELATDGGALLAGRL 3026
Db 3543 RVDLPTYSFRRDRYWLDA--PAADTAVDTAGLGLGTADHPLLGAVVSLPDRDGLLLTGRL 3600
Qy 3027 SLRSHPWLADHAVGGTVLLSGATFLELALHAGTYVGCDRVDELTLHAPLVVPVDGGVSVQ 3086
Db 3601 SLRTHPWLADHAVLGSVLLPGAAMVELAAHAAESAGLRDVRELTLLEPLVLPEHGGVELR 3660
Qy 3087 VGVAAADGE--------GRRLVSVYARGGSACGGGGASGGVWTCHASGVLVEAAAGGVVV 3138
Db 3661 VTVGAPAGEPGGESAGDGARPVSLHSRLADA-----PAGTAWSCHATGLLATDRPELPVA 3715
Qy 3139 DGLAGVWPPRGAVAVDVDGVRDRLAGAGCVLGPVFSGLRAVWRDGGDLLAEVCLP----- 3193
Db 3716 PDRAAMWPPQGAEEVPLDGLYERLDGNGLAFGPLFQGLNAVWRYEGEVFADIALPATTNA 3775
Qy 3194 -----EEAWGDAAG--FGLHPALLDGVVQPLSVLLPGGTGFGEGAGFGEGVRVPAVWGGV 3246
Db 3776 TAPATANGGGSAAAAPYGIHPALLDASLHAIAV---GGL-----VDEPELVRVPFHWSGV 3827
Qy 3247 SLHRAGVTGVRVRVSAVGRGGGREAVSVVVGDEAGVPVASVDRLELRPVDMGQLRAVSVS 3306
Db 3828 TVHAAGAAAARVRLAS----AGTDAVSLSLTDGEGRPLVSVERLTLRPVTADQAAASRVG 3883
Qy 3307 AGRRGSLYAVQW---------------AEVGPVPVCGQ--------------------A 3330
Db 3884 ----GLMHRVAWRPYALASSGEQDPHATSYGPTAVLGKDELKVAAALESAGVEVGLYPDL 3939
Qy 3331 WAWHEDVGESGGGPVPGVVVLRCPDAGAGGGGGGGGGGGVGEVVGGVLGVVQGWLGLERF 3390
Db 3940 AALSQDV--AAGAPAPRTVLAPLP-----AGPADGGAEGVRGTVARTLELLQAWLADEHL 3992
Qy 3391 AGSRLVVVTRGAV----VAGPEDGPVDVVGASVWGLVRSAQAEHPDRFVLLDLDTDTGT- 3445
Db 3993 AGTRLLLVTRGAVRDPEGSGADDGGEDLSHAAAWGLVRTAQTENPGRFGLLDLADDASSY 4052
Qy 3446 ----DLDTGAGAGWGVDGGRVAAVVACGEPQLAVRGERLLAARLKRLESSGDVPAQRSGD 3501
Db 4053 RTLPSVLSDAGLR--------------DEPQLALHDGTIRLARL---------------- 4082
Qy 3502 TRARRSDVPAQRSGGVPARRSVDVSGREVLPWLSGGSVLVTGGTGVLGAAVARHLAGVCG 3561
Db 4083 -----ASVRPETGTAAPA----------LAP---EGTVLLTGGTGGLGGLVARHVVGEWG 4124
Qy 3562 VRDLLLVSRRGPDAPGAEGLRAELAALGAEVRIVACDVGERREVVRLLEGVPAGCPLTGV 3621
Db 4125 VRRLLLVSRRGTDAPGADELVHELEALGADVSVAACDVADREALTAVLDAIPAEHPLTAV 4184
Qy 3622 VHAAGVLDDATIASLTPERLGTVFAAKVDAALLLDELTR--GMELSAFVLFSSAAGILGS 3679
Db 4185 VHTAGVLSDGTLPSMTTEDVEHVLRPKVDAAFLLDELTSTPAYDLAAFVMFSSAAAVFGG 4244
Qy 3680 AGQGNYAAANAALDALAYRRRAAGLPGVSLAWGLWEEASGMTGHLAGTDHRRIIRSGLHP 3739
Db 4245 AGQGAYAAANATLDALAWRRRAAGLPALSLGWGLWAETSGMTGELGQADLRRMSRAGIGG 4304
Qy 3740 MSTPDALALFDAALALDR-PVLLP-----ADLRPA-----PPLPPLLQDLLPATRRR--- 3785
Db 4305 ISDAEGIALLDAALRDDRHPVLLPLRLDAAGLRDAAGNDPAGIPALFRDVVGARTVRARP 4364
Qy 3786 ------TTRTT--TTGGADNGAQLHA-----RLAGQTHEQQHTTLLALVRSHIATVLGHT 3832
Db 4365 SAASASTTAGTAGTPGTADGAAETAAVTLADRAATVDGPARQRLLLEFVVGEVAEVLGHA 4424

```
Qy       3833 TPDTIPPDRAFRDLGFDSLTAVELRNRLSRTTGLRLPTTLAFDHPNPTTLTHHLHTQL-- 3890
Db       4425 RGHRIDAERGFLDLGFDSLTAVELRNRLNSAGGLALPATLVFDHPSPAALASHLDAELPR 4484

Qy       3891 -QPQPDNA-----------------VAPVLAELDKLESALSALDKTDS-ASERVTLRLK 3930
Db       4485 GASDQDGAGNRNGNENGTTASRSTAETDALLAQLTRLEGALVLTGLSDAPGSEEVLEHLR 4544

Qy       3931 SLMLR-------------------------WNAPQHPTAESADDD--EKFTSATEAEIFK 3963
Db       4545 SLRSMVTGETGTGTASGAPDGAGSGAEDRPWAAGDGAGGGSEDGAGVPDFMNASAEELFG 4604

Qy       3964 FIDND 3968
Db       4605 LLDQD 4609


RESULT 3
US-09-988-384B-31
; Sequence 31, Application US/09988384B
; Publication No. US20030073824A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.536US1
;  CURRENT APPLICATION NUMBER: US/09/988,384B
;  CURRENT FILING DATE:  2001-11-19
;  PRIOR APPLICATION NUMBER: PCT/US99/14398
;  PRIOR FILING DATE: 1999-06-25
;  PRIOR APPLICATION NUMBER: US 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 53
; SEQ ID NO 31
;   LENGTH: 4613
;   TYPE: PRT
;   ORGANISM: Streptomyces venezuelae
US-09-988-384B-31

  Query Match             39.1%;  Score 8148;  DB 9;  Length 4613;
  Best Local Similarity   44.8%;  Pred. No. 0;
  Matches 1909;  Conservative  463;  Mismatches 1307;  Indels  586;  Gaps   94;

Qy         15 GEVLGVADEADGGVVFVFPGQGPQWPGMGRELLDASDVFRESVRACEAAFAPYVDWSVEQ 74
Db        620 GLVRGVASGV-GRVAFVFPGQGTQWAGMGAELLDSSAVFAAAMAECEAALSPYVDWSLEA 678

Qy         75 VLRDSPDAPGLDRVDVVQPTLFAVMISLAALWRSQGVEPCAVLGHSLGEIAAAHVSGGLS 134
Db        679 VVRQAPGAPTLERVDVVQPVTFAVMVSLARVWQHHGVTPQAVVGHSQGEIAAAYVAGALS 738

Qy        135 LADAARVVTLWSQA-QTTLAGTGALVSVAATPDELLPRIAPWTEDNPARLAVAAVNGPRS 193
Db        739 LDDAARVVTLRSKSIAAHLAGKGGMLSLALSEDAVLERLAGFD-----GLSVAAVNGPTA 793

Qy        194 TVVSGAREAVADLVADLTAAQVRTRMIPVDVPAHSPLMYAIEERVVSGLLPITPRPSRIP 253
Db        794 TVVSGDPVQIEELARACEADGVRARVIPVDYASHSRQVEIIESELAEVLAGLSPQAPRVP 853

Qy        254 FHSSVTGGRLDTRELDAAYWYRNMSSTVRFEPAAR-LLLQQGPKTFVEMSPHPVLTMGLQ 312
Db        854 FFSTLEGAWITEPVLDGGYWYRNLRHRVGFAPAVETLATDEGFTHFVEVSAHPVLTMALP 913

Qy        313 ELAPDLGDTTGTADTVIMGTLRRGQGTLDHFLTSLAQ---------------LRGHGET 356
Db        914 ------GTVTGLA------TLRRDNGGQDRLVASLAEAWANGLAVDWSPLLPSATGHHSD 961

Qy        357 SATTVLS------ARLTALSPT----------------------QQQSLLLDLVRAHTM 387
Db        962 LPTYAFQTERHWLGEIEALAPAGEPAVQPAVLRTEAAEPAELDRDEQLRVILDKVRAQTA 1021

Qy        388 AVLNDDGNERTASDAGPSASFAHLGFDSVMGVELRNRLSKATGLRLPVTLIFDHTTPAAV 447
Db       1022 QVL----GYATGGQIEVDRTFREAGCTSLTGVDLRNRINAAFGVRMAPSMIFDFPTPEAL 1077

Qy        448 AARLRTAALGHLDEDTAPVPDSPSGHGGTAAADDPIAIIGMACRFPGGVRSPKDLWELAA 507
Db       1078 AEQLLLVVHGEAAANPAGAEPAPVAAAG--AVDEPVAIVGMACRLPGGVASPEDLWRLVA 1135

Qy        508 SGGDAIGPFPTDRGWPTEQRHAQDPTQPGTFYPQGGGFLHDAAHFDAGFFGISPREALAM 567
Db       1136 GGGDAISEFPQDRGWDVEGLYHPDPEHPGTSYVRQGGFIENVAGFDAAFFGISPREALAM 1195

Qy        568 DPQQRLLLETSWEAFERAGIDPLSVRGSRTGVFAGALSFDYGPRMDTASSEGAADVEGHI 627
Db       1196 DPQQRLLLETSWEAVEDAGIDPTSLRGRQVGVFTGAMTHEYGPSL----RDGGEGLDGYL 1251

Qy        628 LTGTTGSVLSGRIAYSFGLEGPAITVDTGCSASLVTLHLACQSLRSGECTLALAGGVSVM 687
Db       1252 LTGNTASVMSGRVSYTLGLEGPALTVDTACSSSLVALHLAVQALRKGEVDMALAGGVAVM 1311

Qy        688 STLGMFIEFSRQRGLSVDGRCKAYSAAADGTGWGEGVGMLLVERLSDAVRLGHRVLAVVR 747
Db       1312 PTPGMFVEFSRQRGLAGDGRSKAFAASADGTSWSEGVGVLLVERLSDARRNGHQVLAVVR 1371

Qy        748 GSAVNQDGASNGLTAPNGPAQERVIRQALANAGLSVADVDVVEGHGTGTTLGDPIEAQAL 807
Db       1372 GSALNQDGASNGLTAPNGPSQQRVIRRALADARLTTSDVDVVEAHGTGTRLGDPIEAQAL 1431

Qy        808 LATYGQ-RAGDRPLWLGSLKSNIGHTMAAAGVGGVIKMVMALREGVLPRTLHVDKPSPQV 866
Db       1432 IATYGQGRDDEQPLRLGSLKSNIGHTQAAAGVSGVIKMVQAMRHGLLPKTLHVDEPSDQI 1491

Qy        867 DWSAGAVRLLTEAVPWPGDAAGRLRRAGVSSFGIGGTNAHVILEEAPAAGGCVAGGGVLE 926
Db       1492 DWSAGAVELLTEAVDWPEKQDGGLRRAAVSSFGISGTNAHVVLEEAPV---------VVE 1542

Qy        927 GAPGLAISVAESVAAPVAVSAPVAESVPVPVPVPVPVPVSARSEAGLRAQAEALRQYVAV 986
Db       1543 GA-----SVVEPSVGGSAVGGGV----------TPWVVSAKSAAALDAQIERLAAF-AS 1585

Qy        987 RPDVSLADVGA--------GLACGRAVLEHRAVVLAADREELVQGLGALAAGEPDRRVTT 1038
Db       1586 RDRTDDADAGAVDAGAVAHVLADGRAQFEHRAVALGAGADDLVQAL-----ADPDGLI-R 1639

Qy       1039 GHAPGGDRGGVVFVFPGQGGQWAGMGVRLLASSPVFARRMQACEEALAPWVDWSVVDILR 1098
Db       1640 GTASG--VGRVAFVFPGQGTQWAGMGAELLDSSAVFAAAMAECEAALSPYVDWSLEAVVR 1697

Qy       1099 RDAGDAVWERADVVQPVLFSVMVSLAALWRSYGIEPDAVLGHSQGEIAAAHVCGALSLKD 1158
Db       1698 QAPGAPTLERVDVVQPVTFAVMVSLARVWQHHGVTPQAVVGHSQGEIAAAYVAGALPLDD 1757

Qy       1159 AAKTVALRSRALAA-VRGRGGMASVPLPAQEVEQLIGERWAGRLWVAAVNGPRSTAVSGD 1217
Db       1758 AARVVTLRSKSIAAHLAGKGGMLSLALNEDAVLERLSD--FDGLSVAAVNGPTATVVSGD 1815

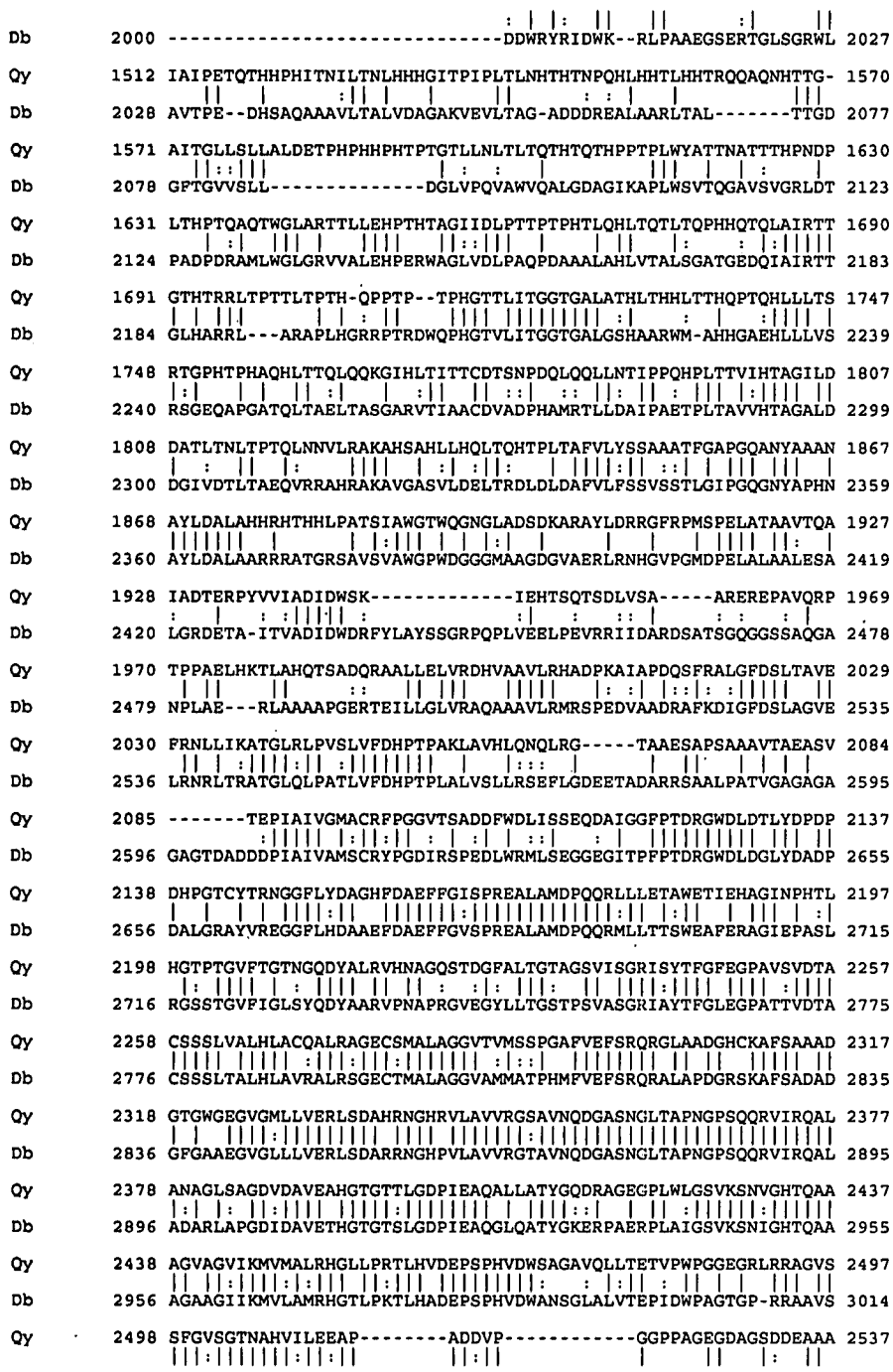
Qy       1218 AEAVDEVLAYCAGTGVRARRIPVDYASHCPHVQPLREELLELLGDISPQPSGVPFFSTVE 1277
Db       1816 PVQIEELAQACKADGFRARIIPVDYASHSRQVEIIESELAQVLAGLSPQAPRVPFFSTLE 1875

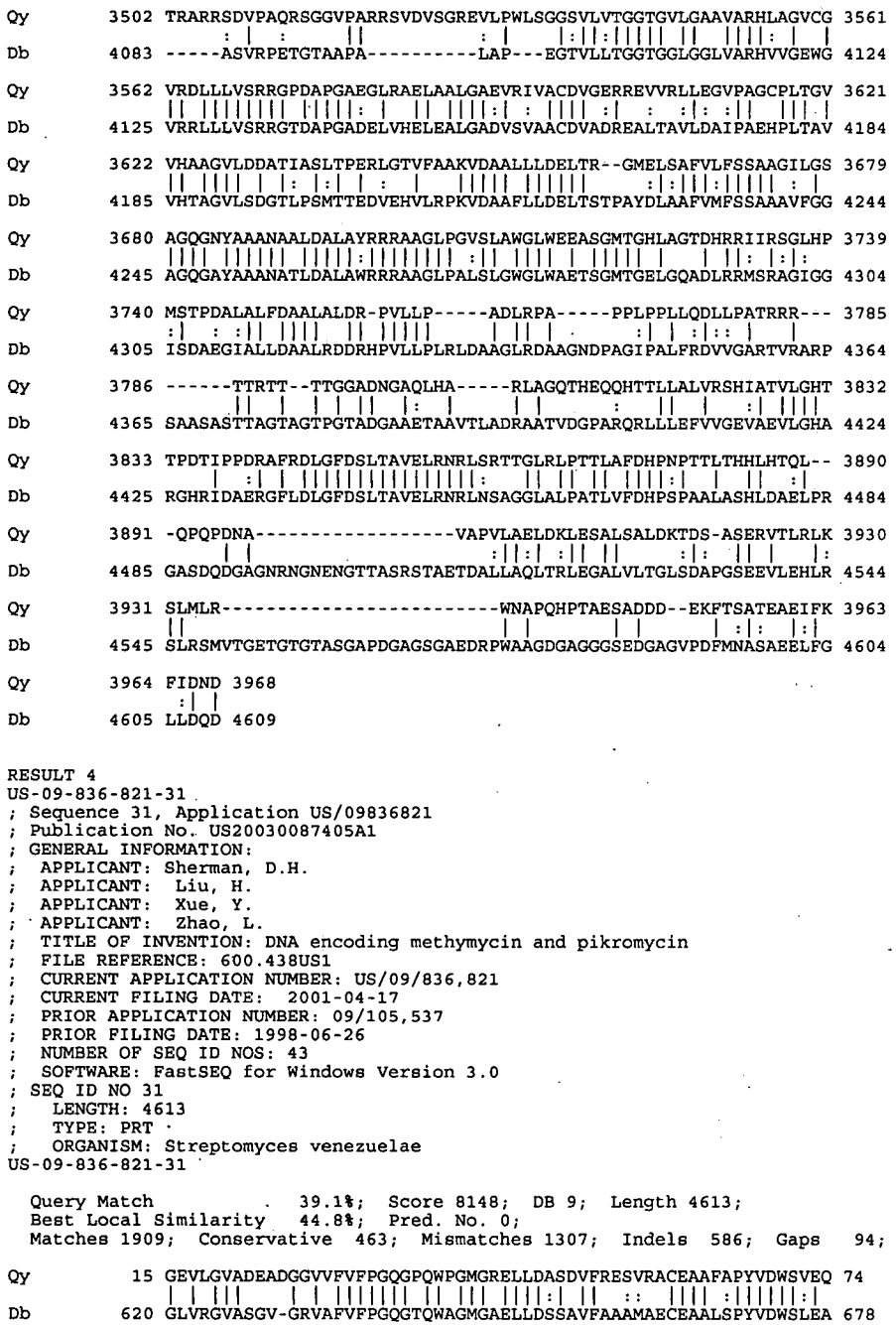
Qy       1278 GTWLDTTTLDAAYWYRNLHQPVRFSDAVQALA-DDGHRVFVEVSPHPTLVPAIEDTTEDT 1336
Db       1876 GTWITEPVLDGTYWYRNLRHRVGFAPAIETLAVDEGFTHFVEVSAHPVLTMTLPET---- 1931

Qy       1337 AEDVTAIGSLRRGDNDTRRFLTALAHTHTTGIGTPTTWHHHYTHHHTHPHPHTHLDLPTY 1396
Db       1932 ---VTGLGTLRREQGGQERLVTSLAEAWVNGL--PVAWTSLLPATASRP------GLPTY 1980

Qy       1397 PFQHQHYWLESSQPGAGSGSGAGAGSGAGSGRAGTAGGTAEVESRFWDAVARQDLETVAT 1456
Db       1981 AFQAERYWLENTPAALATG------------------------------------------ 1999

Qy       1457 TLAVPPSAGLDTVVPALSAWHRHQHDQARINTWTYQETWKPLTLPTTHQPHQT-----WL 1511
```

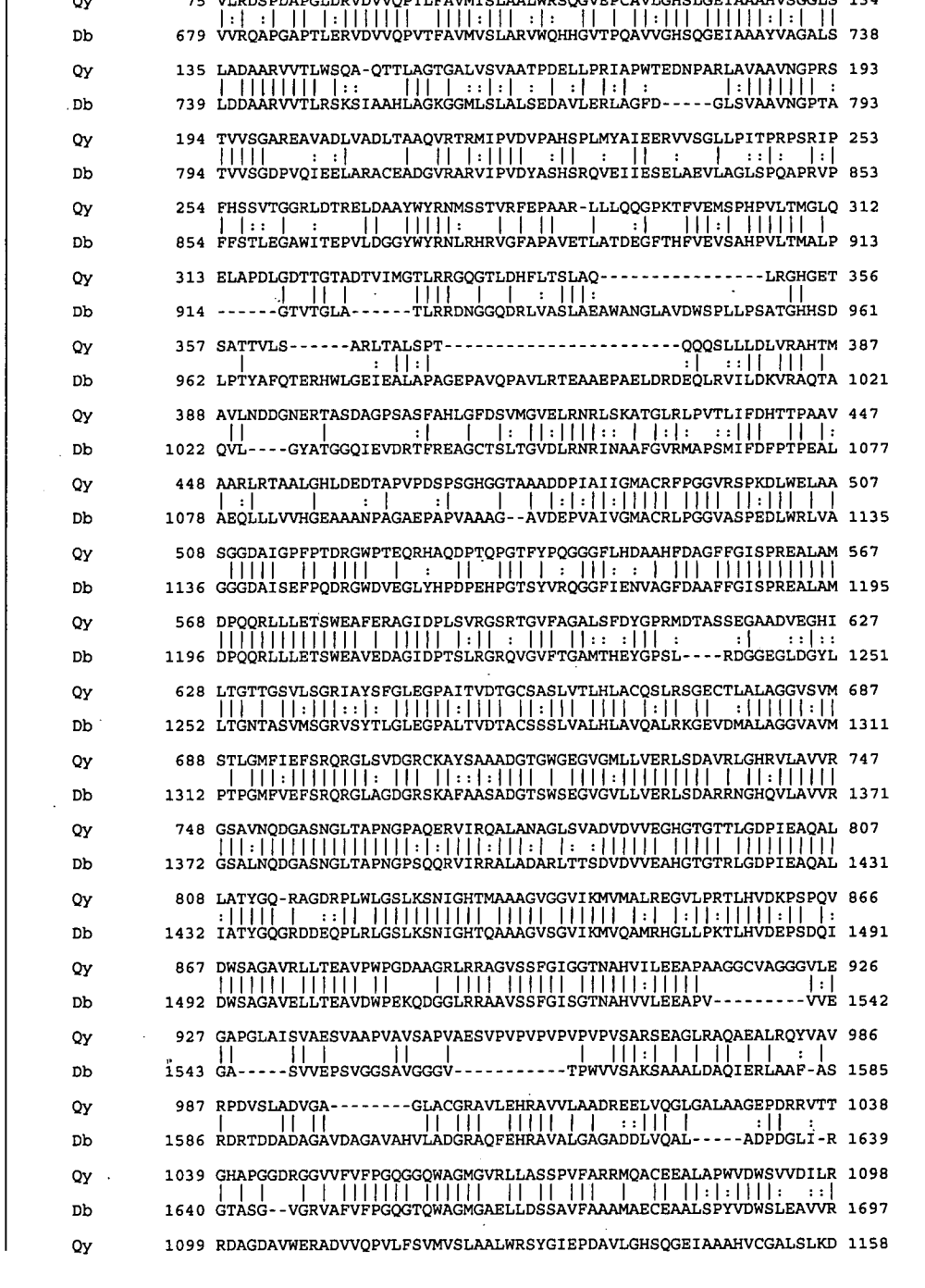

Db 2000 ----------DDWRYRIDWK--RLPAAEGSERTGLSGRWL 2027
Qy 1512 IAIPETQTHHPHITNILTNLHHHGITPIPLTLNHTHTNPQHLHHTLHHTRQQAQNHTTG- 1570
Db 2028 AVTPE--DHSAQAAAVLTALVDAGAKVEVLTAG-ADDDREALAARLTAL-------TTGD 2077
Qy 1571 AITGLLSLLALDETPHPHHPHTPTGTLLNLTLTQTHTQTHPPTPLWYATTNATTTHPNDP 1630
Db 2078 GFTGVVSLL----------DGLVPQVAWVQALGDAGIKAPLWSVTQGAVSVGRLDT 2123
Qy 1631 LTHPTQAQTWGLARTTLLEHPTHTAGIIDLPTTPTPHTLQHLTQTLTQPHHQTQLAIRTT 1690
Db 2124 PADPDRAMLWGLGRVVALEHPERWAGLVDLPAQPDAAALAHLVTALSGATGEDQIAIRTT 2183
Qy 1691 GTHTRRLTPTTLTPTH-QPPTP--TPHGTTLITGGTGALATHLTHHLTTHQPTQHLLLTS 1747
Db 2184 GLHARRL---ARAPLHGRRPTRDWQPHGTVLITGGTGALGSHAARWM-AHHGAEHLLLVS 2239
Qy 1748 RTGPHTPHAQHLTTQLQQKGIHLTITTCDTSNPDQLQQLLNTIPPQHPLTTVIHTAGILD 1807
Db 2240 RSGEQAPGATQLTAELTASGARVTIAACDVADPHAMRTLLDAIPAETPLTAVVHTAGALD 2299
Qy 1808 DATLTNLTPTQLNNVLRAKAHSAHLLHQLTQHTPLTAFVLYSSAAATFGAPGQANYAAAN 1867
Db 2300 DGIVDTLTAEQVRRAHRAKAVGASVLDELTRDLDLDAFVLFSSVSSTLGIPGQGNYAPHN 2359
Qy 1868 AYLDALAHHRHTHHLPATSIAWGTWQGNGLADSDKARAYLDRRGFRPMSPELATAAVTQA 1927
Db 2360 AYLDALAARRRATGRSAVSVAWGPWDGGGMAAGDGVAERLRNHGVPGMDPELALAALESA 2419
Qy 1928 IADTERPYVVIADIDWSK----------IEHTSQTSDLVSA-----ARELREPAVQRP 1969
Db 2420 LGRDETA-ITVADIDWDRFYLAYSSGRPQPLVEELPEVRRIIDARDSATSGQGGSSAQGA 2478
Qy 1970 TPPAELHKTLAHQTSADQRAALLELVRDHVAAVLRHADPKAIAPDQSFRALGFDSLTAVE 2029
Db 2479 NPLAE---RLAAAAPGERTEILLGLVRAQAAAVLRMRSPEDVAADRAFKDIGFDSLAGVE 2535
Qy 2030 FRNLLIKATGLRLPVSLVFDHPTPAKLAVHLQNQLRG-----TAAESAPSAAAVTAEASV 2084
Db 2536 LRNRLTRATGLQLPATLVFDHPTPLALVSLLRSEFLGDEETADARRSAALPATVGAGAGA 2595
Qy 2085 -------TEPIAIVGMACRFPGGVTSADDFWDLISSEQDAIGGFPTDRGWDLDTLYDPDP 2137
Db 2596 GAGTDADDDPIAIVAMSCRYPGDIRSPEDLWRMLSEGGEGITPFPTDRGWDLDGLYDADP 2655
Qy 2138 DHPGTCYTRNGGFLYDAGHFDAEFFGISPREALAMDPQQRLLLETAWETIEHAGINPHTL 2197
Db 2656 DALGRAYVREGGFLHDAAEFDAEFFGVSPREALAMDPQQRMLLTTSWEAFERAGIEPASL 2715
Qy 2198 HGTPTGVFTGTNGQDYALRVHNAGQSTDGFALTGTAGSVISGRISYTFGFEGPAVSVDTA 2257
Db 2716 RGSSTGVFIGLSYQDYAARVPNAPRGVEGYLLTGSTPSVASGRIAYTFGLEGPATTVDTA 2775
Qy 2258 CSSSLVALHLACQALRAGECSMALAGGVTVMSSPGAFVEFSRQRGLAADGHCKAFSAAAD 2317
Db 2776 CSSSLTALHLAVRALRSGECTMALAGGVAMMATPHMFVEFSRQRALAPDGRSKAFSADAD 2835
Qy 2318 GTGWGEGVGMLLVERLSDAHRNGHRVLAVVRGSAVNQDGASNGLTAPNGPSQQRVIRQAL 2377
Db 2836 GFGAAEGVGLLLVERLSDARRNGHPVLAVVRGTAVNQDGASNGLTAPNGPSQQRVIRQAL 2895
Qy 2378 ANAGLSAGDVDAVEAHGTGTTLGDPIEAQALLATYGQDRAGEGPLWLGSVKSNVGHTQAA 2437
Db 2896 ADARLAPGDIDAVETHGTGTSLGDPIEAQGLQATYGKERPAERPLAIGSVKSNIGHTQAA 2955
Qy 2438 AGVAGVIKMVMALRHGLLPRTLHVDEPSPHVDWSAGAVQLLTETVPWPGGEGRLRRAGVS 2497
Db 2956 AGAAGIIKMVLAMRHGTLPKTLHADEPSPHVDWANSGLALVTEPIDWPAGTGP-RRAAVS 3014
Qy 2498 SFGVSGTNAHVILEEAP--------ADDVP-----------GGPPAGEGDAGSDDEAAA 2537
Db 3015 SFGISGTNAHVVLEQAPDAAGEVLGADEVPEVSETVAMAGTAGTSEVAEGSEASEAPAAP 3074
Qy 2538 GS-----PGVWPWLVSAKSQPALRAQAQALHAHLTDHPGLDLA---------DVGYTLAH 2583
Db 3075 GSREASLPGHLPWVLSAKDEQSLRGQAAALHAWLSE-PAADLSDADGPARLRDVGYTLAT 3133
Qy 2584 ARAVFDHRATLIAADRDTFLQALQALAAGEPHPAVIHSSAPGGTGTGEAAGKTAFICSGQ 2643
Db 3134 SRTAFAHRAAVTAADRDGFLDGLATLAQGGTSAHVHLDTARDGT--------TAFLFTGQ 3185
Qy 2644 GTQRPGMAHGLYHTHPVFAAALNDICTHLDPHLDHPLLPLLTQNDNDNEDAAALLQQTRY 2703
Db 3186 GSQRPGAGRELYDRHPVFARALDEICAHLDGHLELPLLDVMFAAEGSAE--AALLDETRY 3243
Qy 2704 AQPALFAFQVALHRLLTDGYHITPHYYAGHSLGEITAAHLAGILTLTDATTLITQRATLM 2763
Db 3244 TQCALFALEVALFRLV-ESWGMRPAALLGHSVGEIAAAHVAGVFSLADAARLVAARGRLM 3302
Qy 2764 QTMPP-GTMTTLHTTPHHITHHLTAHE---NDLAIAAINTPTSLVISGTPHTVQHITTLC 2819
Db 3303 QELPAGGAMLAVQAAEDEIRVWLETEERYAGRLDVAAVNGPEAAVLSGDADAAREAEAYW 3362
Qy 2820 QQQGIKTKTLPTNHAFHSPHTNPILNQLHQHTQTLTYHPPHTPLITANT----PPDQLLT 2875
Db 3363 SGLGRRTRALRVSHAFHSAHMDGMLDGFRAVLETVEFRRPSLTVVSNVTGLAAGPDDLCD 3422
Qy 2876 PHYWTQQARNTVDYATTTQTLHQHGVTTYIELGPDNTLTTLTHHNLPNPPTTTLT---LT 2932
Db 3423 PEYWVRHVRGTVRFLDGVRVLRDLGVRTCLELGPDGVLTAMAADGLADTPADSAAGSPVG 3482
Qy 2933 HPHHHPQTH---------LLTNL-----AKTTTTWHPHHYTH-HDNQPHTHT-------- 2969
Db 3483 SPAGSPADSAAGALRPRPLLVALLRRKRSETETVADALGRAHAHGTGPDWHAWFAGSGAH 3542
Qy 2970 HLDLPTYPFQHHHYWLESTQPGAG---NVSAAGLDPTEHPLLGATLELATDGGALLAGRL 3026
Db 3543 RVDLPTYSFRRDRYWLDA--PAADTAVDTAGLGLGTADHPLLGAVVSLPDRDGLLLTGRL 3600
Qy 3027 SLRSHPWLADHAVGGTVLLSGATFLELALHAGTYVGCDRVDELTLHAPLVVPVDGGVSVQ 3086
Db 3601 SLRTHPWLADHAVLGSVLLPGAAMVELAAHAAESAGLRDVRELTLLEPLVLPEHGGVELR 3660
Qy 3087 VGVAAADGE--------GRRLVSVYARGGSACGGGGASGGVWTCHASGVLVEAAAGGVVV 3138
Db 3661 VTVGAPAGEPGGESAGDGARPVSLHSRLADA-----PAGTAWSCHATGLLATDRPELPVA 3715
Qy 3139 DGLAGVWPPRGAVAVDVDGVRDRLAGAGCVLGPVFSGLRAVWRDGGDLLAEVCLP----- 3193
Db 3716 PDRAAMWPPQGAEEVPLDGLYERLDGNGLAFGPLFQGLNAVWRYEGEVFADIALPATTNA 3775
Qy 3194 -----EEAWGDAAG--FGLHPALLDGVVQPLSVLLPGGTGFGEGAGFGEGVRVPAVWGGV 3246
Db 3776 TAPATANGGGSAAAAPYGIHPALLDASLHAIAV---GGL-----VDEPELVRVPFHWSGV 3827
Qy 3247 SLHRAGVTGVRVRVSAVGRGGGREAVSVVVGDEAGVPVASVDRLELRPVDMGQLRAVSVS 3306
Db 3828 TVHAAGAAAARVRLAS----AGTDAVSLSLTDGEGRPLVSVERLTLRPVTADQAAASRVG 3883
Qy 3307 AGRRGSLYAVQW---------------AEVGPVPVCGQ--------------------A 3330
Db 3884 ----GLMHRVAWRPYALASSGEQDPHATSYGPTAVLGKDELKVAAALESAGVEVGLYPDL 3939
Qy 3331 WAWHEDVGESGGGPVPGVVVLRCPDAGAGGGGGGGGGGVGEVVGGVLGVVQGWLGLERF 3390
Db 3940 AALSQDV--AAGAPAPRTVLAPLP-----AGPADGGAEGVRGTVARTLELLQAWLADEHL 3992
Qy 3391 AGSRLVVVTRGAV----VAGPEDGPVDVVGASVWGLVRSAQAEHPDRFVLLDLDTDTGT- 3445
Db 3993 AGTRLLLVTRGAVRDPEGSGADDGGEDLSHAAAWGLVRTAQTENPGRFGLLDLADDASSY 4052
Qy 3446 ----DLDTGAGAGWGVDGGRVAAVVACGEPQLAVRGERLLAARLKRLESSGDVPAQRSGD 3501
Db 4053 RTLPSVLSDAGLR--------------DEPQLALHDGTIRLARL---------------- 4082

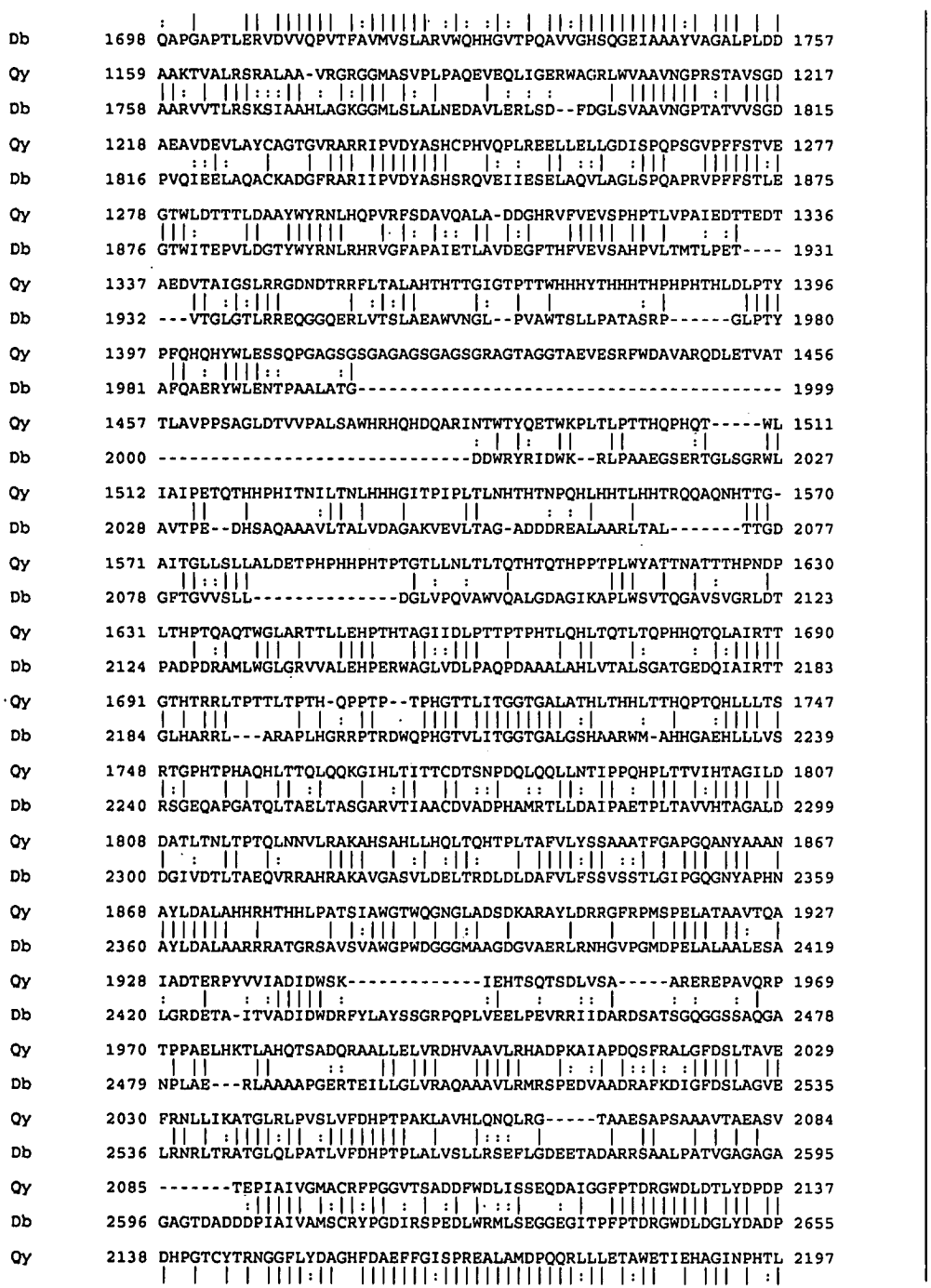


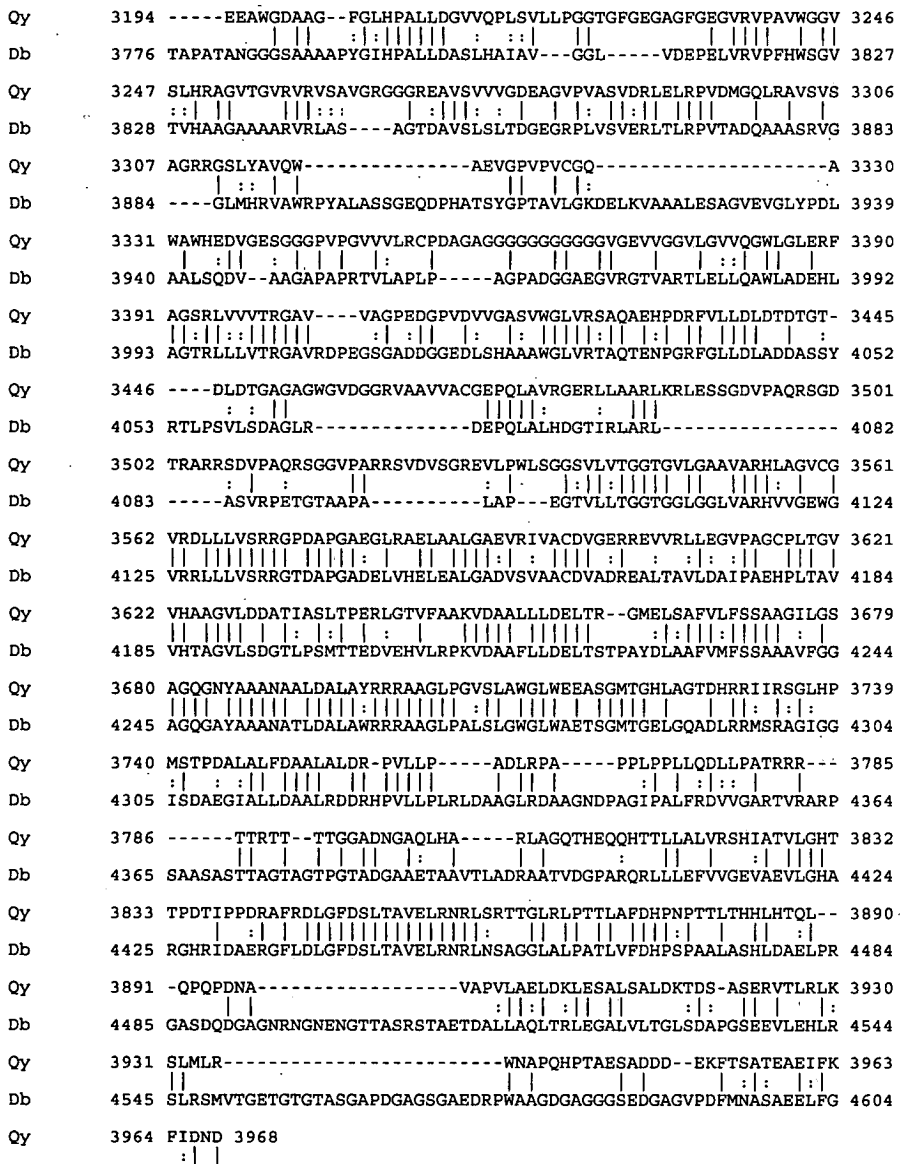
RESULT 4
US-09-836-821-31
; Sequence 31, Application US/09836821
; Publication No. US20030087405A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/836,821
; CURRENT FILING DATE: 2001-04-17
; PRIOR APPLICATION NUMBER: 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 31
; LENGTH: 4613
; TYPE: PRT
; ORGANISM: Streptomyces venezuelae
US-09-836-821-31

Query Match 39.1%; Score 8148; DB 9; Length 4613;
Best Local Similarity 44.8%; Pred. No. 0;
Matches 1909; Conservative 463; Mismatches 1307; Indels 586; Gaps 94;



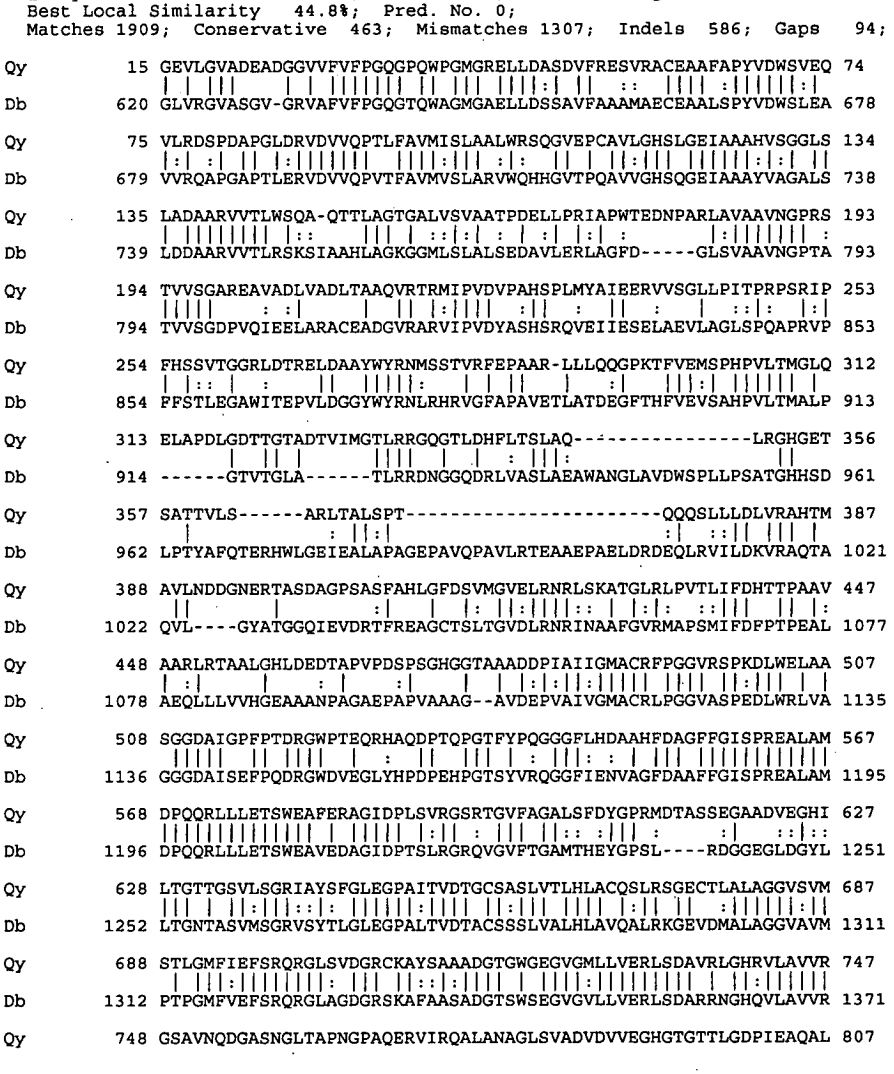


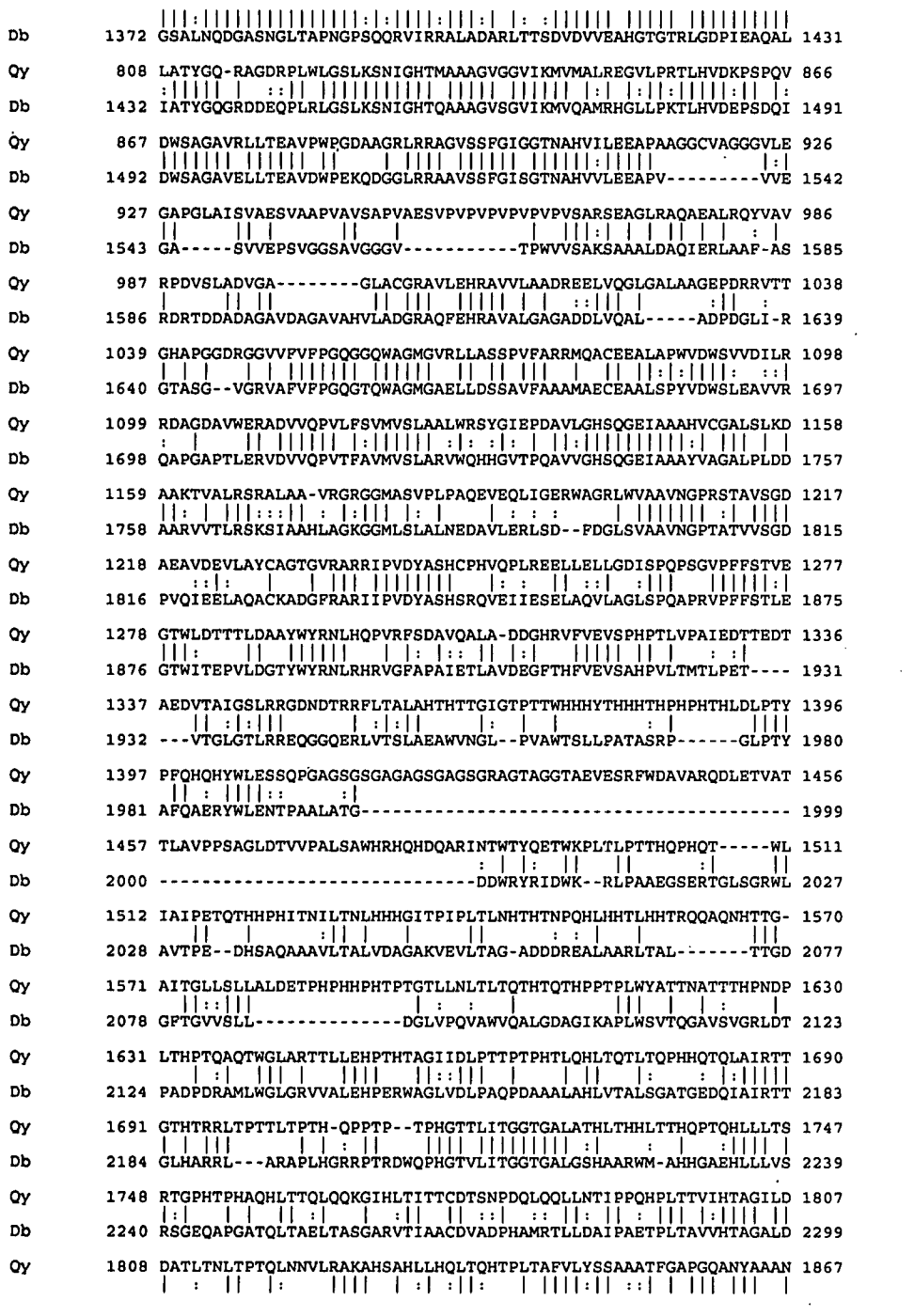

Db 1698 QAPGAPTLERVDVVQPVTFAVMVSLARVWQHHGVTPQAVVGHSQGEIAAAYVAGALPLDD 1757
Qy 1159 AAKTVALRSRALAA-VRGRGGMASVPLPAQEVEQLIGERWAGRLWVAAVNGPRSTAVSGD 1217
Db 1758 AARVVTLRSKSIAAHLAGKGGMLSLALNEDAVLERLSD--FDGLSVAAVNGPTATVVSGD 1815
Qy 1218 AEAVDEVLAYCAGTGVRARRIPVDYASHCPHVQPLREELLELLGDISPQPSGVPFFSTVE 1277
Db 1816 PVQIEELAQACKADGFRARIIPVDYASHSRQVEIIESELAQVLAGLSPQAPRVPFFSTLE 1875
Qy 1278 GTWLDTTTLDAAYWYRNLHQPVRFSDAVQALA-DDGHRVFVEVSPHPTLVPAIEDTTEDT 1336
Db 1876 GTWITEPVLDGTYWYRNLRHRVGFAPAIETLAVDEGFTHFVEVSAHPVLTMTLPET---- 1931
Qy 1337 AEDVTAIGSLRRGDNDTRRFLTALAHTHTTGIGTPTTWHHHYTHHHTHPHPHTHLDLPTY 1396
Db 1932 ---VTGLGTLRREQGGQERLVTSLAEAWVNGL--PVAWTSLLPATASRP------GLPTY 1980
Qy 1397 PFQHQHYWLESSQPGAGSGSGAGAGSGAGSGRAGTAGGTAEVESRFWDAVARQDLETVAT 1456
Db 1981 AFQAERYWLENTPAALATG---------- 1999
Qy 1457 TLAVPPSAGLDTVVPALSAWHRHQHDQARINTWTYQETWKPLTLPTTHQPHQT-----WL 1511
Db 2000 ----------DDWRYRIDWK--RLPAAEGSERTGLSGRWL 2027
Qy 1512 IAIPETQTHHPHITNILTNLHHHGITPIPLTLNHTHTNPQHLHHTLHHTRQQAQNHTTG- 1570
Db 2028 AVTPE--DHSAQAAAVLTALVDAGAKVEVLTAG-ADDDREALAARLTAL-------TTGD 2077
Qy 1571 AITGLLSLLALDETPHPHHPHTPTGTLLNLTLTQTHTQTHPPTPLWYATTNATTTHPNDP 1630
Db 2078 GFTGVVSLL--------------DGLVPQVAWVQALGDAGIKAPLWSVTQGAVSVGRLDT 2123
Qy 1631 LTHPTQAQTWGLARTTLLEHPTHTAGIIDLPTTPTPHTLQHLTQTLTQPHHQTQLAIRTT 1690
Db 2124 PADPDRAMLWGLGRVVALEHPERWAGLVDLPAQPDAAALAHLVTALSGATGEDQIAIRTT 2183
Qy 1691 GTHTRRLTPTTLTPTH-QPPTP--TPHGTTLITGGTGALATHLTHHLTTHQPTQHLLLTS 1747
Db 2184 GLHARRL---ARAPLHGRRPTRDWQPHGTVLITGGTGALGSHAARWM-AHHGAEHLLLVS 2239
Qy 1748 RTGPHTPHAQHLTTQLQQKGIHLTITTCDTSNPDQLQQLLNTIPPQHPLTTVIHTAGILD 1807
Db 2240 RSGEQAPGATQLTAELTASGARVTIAACDVADPHAMRTLLDAIPAETPLTAVVHTAGALD 2299
Qy 1808 DATLTNLTPTQLNNVLRAKAHSAHLLHQLTQHTPLTAFVLYSSAAATFGAPGQANYAAAN 1867
Db 2300 DGIVDTLTAEQVRRAHRAKAVGASVLDELTRDLDLDAFVLFSSVSSTLGIPGQGNYAPHN 2359
Qy 1868 AYLDALAHHRHTHHLPATSIAWGTWQGNGLADSDKARAYLDRRGFRPMSPELATAAVTQA 1927
Db 2360 AYLDALAARRRATGRSAVSVAWGPWDGGGMAAGDGVAERLRNHGVPGMDPELALAALESA 2419
Qy 1928 IADTERPYVVIADIDWSK------------IEHTSQTSDLVSA-----ARELREPAVQRP 1969
Db 2420 LGRDETA-ITVADIDWDRFYLAYSSGRPQPLVEELPEVRRIIDARDSATSGQGGSSAQGA 2478
Qy 1970 TPPAELHKTLAHQTSADQRAALLELVRDHVAAVLRHADPKAIAPDQSFRALGFDSLTAVE 2029
Db 2479 NPLAE---RLAAAAPGERTEILLGLVRAQAAAVLRMRSPEDVAADRAFKDIGFDSLAGVE 2535
Qy 2030 FRNLLIKATGLRLPVSLVFDHPTPAKLAVHLQNQLRG-----TAAESAPSAAAVTAEASV 2084
Db 2536 LRNRLTRATGLQLPATLVFDHPTPLALVSLLRSEFLGDEETADARRSAALPATVGAGAGA 2595
Qy 2085 -------TEPIAIVGMACRFPGGVTSADDFWDLISSEQDAIGGFPTDRGWDLDTLYDPDP 2137
Db 2596 GAGTDADDDPIAIVAMSCRYPGDIRSPEDLWRMLSEGGEGITPFPTDRGWDLDGLYDADP 2655
Qy 2138 DHPGTCYTRNGGFLYDAGHFDAEFFGISPREALAMDPQQRLLLETAWETIEHAGINPHTL 2197
Db 2656 DALGRAYVREGGFLHDAAEFDAEFFGVSPREALAMDPQQRMLLTTSWEAFERAGIEPASL 2715
Qy 2198 HGTPTGVFTGTNGQDYALRVHNAGQSTDGFALTGTAGSVISGRISYTFGFEGPAVSVDTA 2257
Db 2716 RGSSTGVFIGLSYQDYAARVPNAPRGVEGYLLTGSTPSVASGRIAYTFGLEGPATTVDTA 2775
Qy 2258 CSSSLVALHLACQALRAGECSMALAGGVTVMSSPGAFVEFSRQRGLAADGHCKAFSAAAD 2317
Db 2776 CSSSLTALHLAVRALRSGECTMALAGGVAMMATPHMFVEFSRQRALAPDGRSKAFSADAD 2835
Qy 2318 GTGWGEGVGMLLVERLSDAHRNGHRVLAVVRGSAVNQDGASNGLTAPNGPSQQRVIRQAL 2377
Db 2836 GFGAAEGVGLLLVERLSDARRNGHPVLAVVRGTAVNQDGASNGLTAPNGPSQQRVIRQAL 2895
Qy 2378 ANAGLSAGDVDAVEAHGTGTTLGDPIEAQALLATYGQDRAGEGPLWLGSVKSNVGHTQAA 2437
Db 2896 ADARLAPGDIDAVETHGTGTSLGDPIEAQGLQATYGKERPAERPLAIGSVKSNIGHTQAA 2955
Qy 2438 AGVAGVIKMVMALRHGLLPRTLHVDEPSPHVDWSAGAVQLLTETVPWPGGEGRLRRAGVS 2497
Db 2956 AGAAGIIKMVLAMRHGTLPKTLHADEPSPHVDWANSGLALVTEPIDWPAGTGP-RRAAVS 3014
Qy 2498 SFGVSGTNAHVILEEAP--------ADDVP-----------GGPPAGEGDAGSDDEAAA 2537
Db 3015 SFGISGTNAHVVLEQAPDAAGEVLGADEVPEVSETVAMAGTAGTSEVAEGSEASEAPAAP 3074
Qy 2538 GS-----PGVWPWLVSAKSQPALRAQAQALHAHLTDHPGLDLA---------DVGYTLAH 2583
Db 3075 GSREASLPGHLPWVLSAKDEQSLRGQAAALHAWLSE-PAADLSDADGPARLRDVGYTLAT 3133
Qy 2584 ARAVFDHRATLIAADRDTFLQALQALAAGEPHPAVIHSSAPGGTGTGEAAGKTAFICSGQ 2643
Db 3134 SRTAFAHRAAVTAADRDGFLDGLATLAQGGTSAHVHLDTARDGT--------TAFLFTGQ 3185
Qy 2644 GTQRPGMAHGLYHTHPVFAAALNDICTHLDPHLDHPLLPLLTQNDNDNEDAAALLQQTRY 2703
Db 3186 GSQRPGAGRELYDRHPVFARALDEICAHLDGHLELPLLDVMFAAEGSAE--AALLDETRY 3243
Qy 2704 AQPALFAFQVALHRLLTDGYHITPHYYAGHSLGEITAAHLAGILTLTDATTLITQRATLM 2763
Db 3244 TQCALFALEVALFRLV-ESWGMRPAALLGHSVGEIAAAHVAGVFSLADAARLVAARGRLM 3302
Qy 2764 QTMPP-GTMTTLHTTPHHITHHLTAHE---NDLAIAAINTPTSLVISGTPHTVQHITTLC 2819
Db 3303 QELPAGGAMLAVQAAEDEIRVWLETEERYAGRLDVAAVNGPEAAVLSGDADAAREAEAYW 3362
Qy 2820 QQQGIKTKTLPTNHAFHSPHTNPILNQLHQHTQTLTYHPPHTPLITANT----PPDQLLT 2875
Db 3363 SGLGRRTRALRVSHAFHSAHMDGMLDGFRAVLETVEFRRPSLTVVSNVTGLAAGPDDLCD 3422
Qy 2876 PHYWTQQARNTVDYATTTQTLHQHGVTTYIELGPDNTLTTLTHHNLPNPPTTTLT---LT 2932
Db 3423 PEYWVRHVRGTVRFLDGVRVLRDLGVRTCLELGPDGVLTAMAADGLADTPADSAAGSPVG 3482
Qy 2933 HPHHHPQTH---------LLTNL-----AKTTTTWHPHHYTH-HDNQPHTHT-------- 2969
Db 3483 SPAGSPADSAAGALRPRPLLVALLRRKRSETETVADALGRAHAHGTGPDWHAWFAGSGAH 3542
Qy 2970 HLDLPTYPFQHHHYWLESTQPGAG---NVSAAGLDPTEHPLLGATLELATDGGALLAGRL 3026
Db 3543 RVDLPTYSFRRDRYWLDA--PAADTAVDTAGLGLGTADHPLLGAVVSLPDRDGLLLTGRL 3600
Qy 3027 SLRSHPWLADHAVGGTVLLSGATFLELALHAGTYVGCDRVDELTLHAPLVVPVDGGVSVQ 3086
Db 3601 SLRTHPWLADHAVLGSVLLPGAAMVELAAHAAESAGLRDVRELTLLEPLVLPEHGGVELR 3660
Qy 3087 VGVAAADGE--------GRRLVSVYARGGSACGGGGASGGVWTCHASGVLVEAAAGGVVV 3138
Db 3661 VTVGAPAGEPGGESAGDGARPVSLHSRLADA-----PAGTAWSCHATGLLATDRPELPVA 3715
Qy 3139 DGLAGVWPPRGAVAVDVDGVRDRLAGAGCVLGPVPSGLRAVWRDGGDLLAEVCLP----- 3193
Db 3716 PDRAAMWPPQGAEEVPLDGLYERLDGNGLAFGPLPQGLNAVWRYEGEVFADIALPATTNA 3775

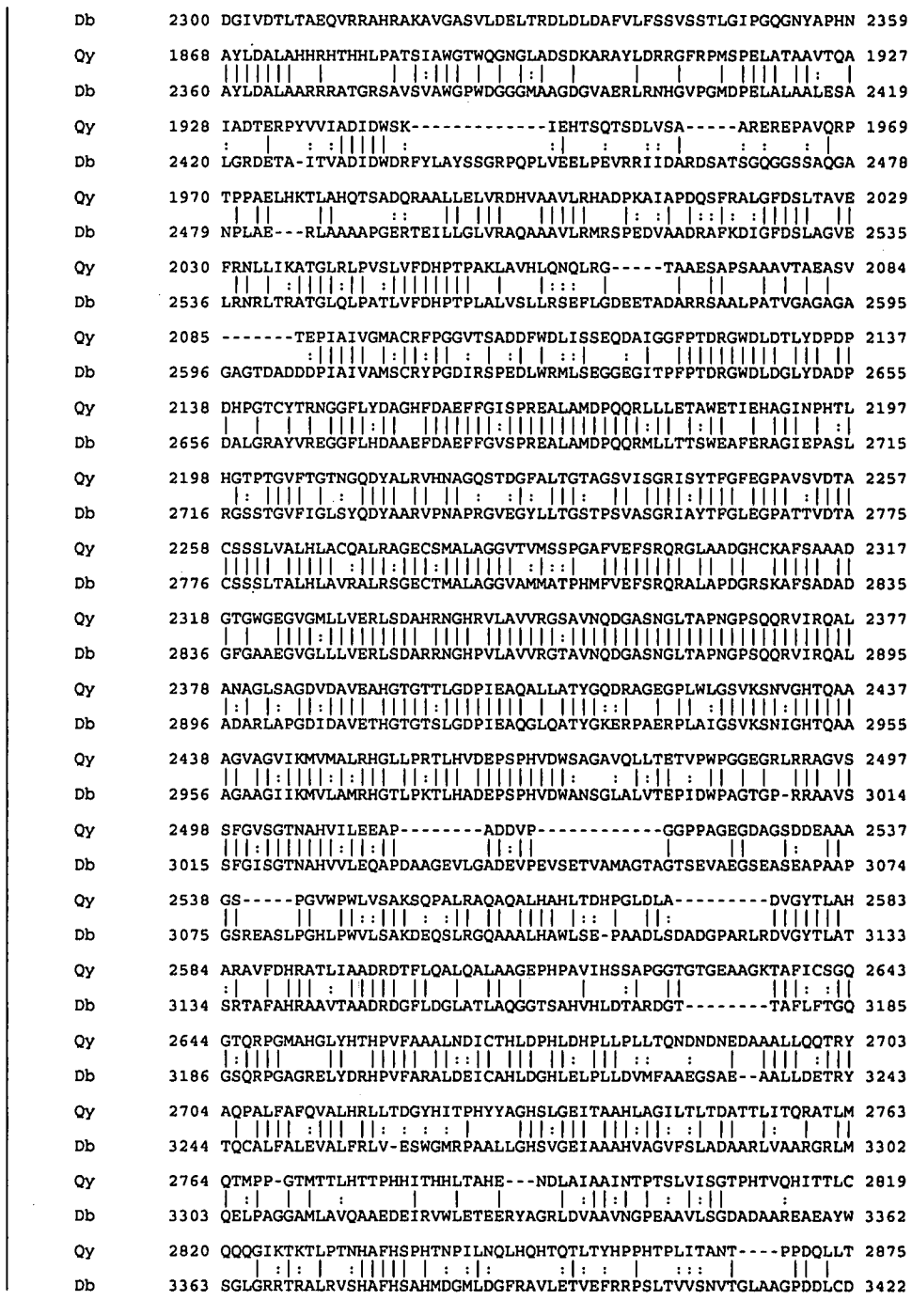


RESULT 5
US-09-861-289-31
; Sequence 31, Application US/09861289
; Patent No. US20020110897A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/861,289
; CURRENT FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 31
; LENGTH: 4613
; TYPE: PRT
; ORGANISM: Streptomyces venezuelae
US-09-861-289-31

Query Match 39.1%; Score 8148; DB 10; Length 4613;
Best Local Similarity 44.8%; Pred. No. 0;
Matches 1909; Conservative 463; Mismatches 1307; Indels 586; Gaps 94;




Db 1372 GSALNQDGASNGLTAPNGPSQQRVIRRALADARLTTSDVDVVEAHGTGTRLGDPIEAQAL 1431
Qy 808 LATYGQ-RAGDRPLWLGSLKSNIGHTMAAAGVGGVIKMVMALREGVLPRTLHVDKPSPQV 866
Db 1432 IATYGQGRDDEQPLRLGSLKSNIGHTQAAAGVSGVIKMVQAMRHGLLPKTLHVDEPSDQI 1491
Qy 867 DWSAGAVRLLTEAVPWPGDAAGRLRRAGVSSFGIGGTNAHVILEEAPAAGGCVAGGGVLE 926
Db 1492 DWSAGAVELLTEAVDWPEKQDGGLRRAAVSSFGISGTNAHVVLEEAPV---------VVE 1542
Qy 927 GAPGLAISVAESVAAPVAVSAPVAESVPVPVPVPVPVPVSARSEAGLRAQAEALRQYVAV 986
Db 1543 GA-----SVVEPSVGGSAVGGGV-----------TPWVVSAKSAAALDAQIERLAAF-AS 1585
Qy 987 RPDVSLADVGA--------GLACGRAVLEHRAVVLAADREELVQGLGALAAGEPDRRVTT 1038
Db 1586 RDRTDDADAGAVDAGAVAHVLADGRAQFEHRAVALGAGADDLVQAL-----ADPDGLI-R 1639
Qy 1039 GHAPGGDRGGVVFVFPGQGGQWAGMGVRLLASSPVFARRMQACEEALAPWVDWSVVDILR 1098
Db 1640 GTASG--VGRVAFVFPGQGTQWAGMGAELLDSSAVFAAAMAECEAALSPYVDWSLEAVVR 1697
Qy 1099 RDAGDAVWERADVVQPVLFSVMVSLAALWRSYGIEPDAVLGHSQGEIAAAHVCGALSLKD 1158
Db 1698 QAPGAPTLERVDVVQPVTFAVMVSLARVWQHHGVTPQAVVGHSQGEIAAAYVAGALPLDD 1757
Qy 1159 AAKTVALRSRALAA-VRGRGGMASVPLPAQEVEQLIGERWAGRLWVAAVNGPRSTAVSGD 1217
Db 1758 AARVVTLRSKSIAAHLAGKGGMLSLALNEDAVLERLSD--FDGLSVAAVNGPTATVVSGD 1815
Qy 1218 AEAVDEVLAYCAGTGVRARRIPVDYASHCPHVQPLREELLELLGDISPQPSGVPFFSTVE 1277
Db 1816 PVQIEELAQACKADGFRARIIPVDYASHSRQVEIIESELAQVLAGLSPQAPRVPFFSTLE 1875
Qy 1278 GTWLDTTTLDAAYWYRNLHQPVRFSDAVQALA-DDGHRVFVEVSPHPTLVPAIEDTTEDT 1336
Db 1876 GTWITEPVLDGTYWYRNLRHRVGFAPAIETLAVDEGFTHFVEVSAHPVLTMTLPET---- 1931
Qy 1337 AEDVTAIGSLRRGDNDTRRFLTALAHTHTTGIGTPTTWHHHYTHHHTHPHPHTHLDLPTY 1396
Db 1932 ---VTGLGTLRREQGGQERLVTSLAEAWVNGL--PVAWTSLLPATASRP------GLPTY 1980
Qy 1397 PFQHQHYWLESSQPGAGSGSGAGAGSGAGSGRAGTAGGTAEVESRFWDAVARQDLETVAT 1456
Db 1981 AFQAERYWLENTPAALATG-- 1999
Qy 1457 TLAVPPSAGLDTVVPALSAWHRHQHDQARINTWTYQETWKPLTLPTTHQPHQT-----WL 1511
Db 2000 ----------------------------DDWRYRIDWK--RLPAAEGSERTGLSGRWL 2027
Qy 1512 IAIPETQTHHPHITNILTNLHHHGITPIPLTLNHTHTNPQHLHHTLHHTRQQAQNHTTG- 1570
Db 2028 AVTPE--DHSAQAAAVLTALVDAGAKVEVLTAG-ADDDREALAARLTAL-------TTGD 2077
Qy 1571 AITGLLSLLALDETPHPHHPHTPTGTLLNLTLTQTHTQTHPPTPLWYATTNATTTHPNDP 1630
Db 2078 GPTGVVSLL--------------DGLVPQVAWVQALGDAGIKAPLWSVTQGAVSVGRLDT 2123
Qy 1631 LTHPTQAQTWGLARTTLLEHPTHTAGIIDLPTTPTPHTLQHLTQTLTQPHHQTQLAIRTT 1690
Db 2124 PADPDRAMLWGLGRVVALEHPERWAGLVDLPAQPDAAALAHLVTALSGATGEDQIAIRTT 2183
Qy 1691 GTHTRRLTPTTLTPTH-QPPTP--TPHGTTLITGGTGALATHLTHHLTTHQPTQHLLLTS 1747
Db 2184 GLHARRL---ARAPLHGRRPTRDWQPHGTVLITGGTGALGSHAARWM-AHHGAEHLLLVS 2239
Qy 1748 RTGPHTPHAQHLTTQLQQKGIHLTITTCDTSNPDQLQQLLNTIPPQHPLTTVIHTAGILD 1807
Db 2240 RSGEQAPGATQLTAELTASGARVTIAACDVADPHAMRTLLDAIPAETPLTAVVHTAGALD 2299
Qy 1808 DATLTNLTPTQLNNVLRAKAHSAHLLHQLTQHTPLTAFVLYSSAAATFGAPGQANYAAAN 1867


Db 2300 DGIVDTLTAEQVRRAHRAKAVGASVLDELTRDLDLDAFVLFSSVSSTLGIPGQGNYAPHN 2359
Qy 1868 AYLDALAHHRHTHHLPATSIAWGTWQGNGLADSDKARAYLDRRGFRPMSPELATAAVTQA 1927
Db 2360 AYLDALAARRRATGRSAVSVAWGPWDGGGMAAGDGVAERLRNHGVPGMDPELALAALESA 2419
Qy 1928 IADTERPYVVIADIDWSK------------IEHTSQTSDLVSA-----ARE REPAVQRP 1969
Db 2420 LGRDETA-ITVADIDWDRFYLAYSSGRPQPLVEELPEVRRIIDARDSATSGQGGSSAQGA 2478
Qy 1970 TPPAELHKTLAHQTSADQRAALLELVRDHVAAVLRHADPKAIAPDQSFRALGFDSLTAVE 2029
Db 2479 NPLAE---RLAAAAPGERTEILLGLVRAQAAAVLRMRSPEDVAADRAFKDIGFDSLAGVE 2535
Qy 2030 FRNLLIKATGLRLPVSLVFDHPTPAKLAVHLQNQLRG-----TAAESAPSAAAVTAEASV 2084
Db 2536 LRNRLTRATGLQLPATLVFDHPTPLALVSLLRSEFLGDEETADARRSAALPATVGAGAGA 2595
Qy 2085 -------TEPIAIVGMACRFPGGVTSADDFWDLISSEQDAIGGFPTDRGWDLDTLYDPDP 2137
Db 2596 GAGTDADDDPIAIVAMSCRYPGDIRSPEDLWRMLSEGGEGITPFPTDRGWDLDGLYDADP 2655
Qy 2138 DHPGTCYTRNGGFLYDAGHFDAEFFGISPREALAMDPQQRLLLETAWETIEHAGINPHTL 2197
Db 2656 DALGRAYVREGGFLHDAAEFDAEFFGVSPREALAMDPQQRMLLTTSWEAFERAGIEPASL 2715
Qy 2198 HGTPTGVFTGTNGQDYALRVHNAGQSTDGFALTGTAGSVISGRISYTFGFEGPAVSVDTA 2257
Db 2716 RGSSTGVFIGLSYQDYAARVPNAPRGVEGYLLTGSTPSVASGRIAYTFGLEGPATTVDTA 2775
Qy 2258 CSSSLVALHLACQALRAGECSMALAGGVTVMSSPGAFVEFSRQRGLAADGHCKAFSAAAD 2317
Db 2776 CSSSLTALHLAVRALRSGECTMALAGGVAMMATPHMFVEFSRQRALAPDGRSKAFSADAD 2835
Qy 2318 GTGWGEGVGMLLVERLSDAHRNGHRVLAVVRGSAVNQDGASNGLTAPNGPSQQRVIRQAL 2377
Db 2836 GFGAAEGVGLLLVERLSDARRNGHPVLAVVRGTAVNQDGASNGLTAPNGPSQQRVIRQAL 2895
Qy 2378 ANAGLSAGDVDAVEAHGTGTTLGDPIEAQALLATYGQDRAGEGPLWLGSVKSNVGHTQAA 2437
Db 2896 ADARLAPGDIDAVETHGTGTSLGDPIEAQGLQATYGKERPAERPLAIGSVKSNIGHTQAA 2955
Qy 2438 AGVAGVIKMVMALRHGLLPRTLHVDEPSPHVDWSAGAVQLLTETVPWPGGEGRLRRAGVS 2497
Db 2956 AGAAGIIKMVLAMRHGTLPKTLHADEPSPHVDWANSGLALVTEPIDWPAGTGP-RRAAVS 3014
Qy 2498 SFGVSGTNAHVILEEAP--------ADDVP-----------GGPPAGEGDAGSDDEAAA 2537
Db 3015 SFGISGTNAHVVLEQAPDAAGEVLGADEVPEVSETVAMAGTAGTSEVAEGSEASEAPAAP 3074
Qy 2538 GS-----PGVWPWLVSAKSQPALRAQAQALHAHLTDHPGLDLA---------DVGYTLAH 2583
Db 3075 GSREASLPGHLPWVLSAKDEQSLRGQAAALHAWLSE-PAADLSDADGPARLRDVGYTLAT 3133
Qy 2584 ARAVFDHRATLIAADRDTFLQALQALAAGEPHPAVIHSSAPGGTGTGEAAGKTAFICSGQ 2643
Db 3134 SRTAFAHRAAVTAADRDGFLDGLATLAQGGTSAHVHLDTARDGT--------TAFLFTGQ 3185
Qy 2644 GTQRPGMAHGLYHTHPVFAAALNDICTHLDPHLDHPLLPLLTQNDNDNEDAAALLQQTRY 2703
Db 3186 GSQRPGAGRELYDRHPVFARALDEICAHLDGHLELPLLDVMFAAEGSAE--AALLDETRY 3243
Qy 2704 AQPALFAFQVALHRLLTDGYHITPHYYAGHSLGEITAAHLAGILTLTDATTLITQRATLM 2763
Db 3244 TQCALFALEVALFRLV-ESWGMRPAALLGHSVGEIAAAHVAGVFSLADAARLVAARGRLM 3302
Qy 2764 QTMPP-GTMTTLHTTPHHITHHLTAHE---NDLAIAAINTPTSLVISGTPHTVQHITTLC 2819
Db 3303 QELPAGGAMLAVQAAEDEIRVWLETEERYAGRLDVAAVNGPEAAVLSGDADAAREAEAYW 3362
Qy 2820 QQQGIKTKTLPTNHAFHSPHTNPILNQLHQHTQTLTYHPPHTPLITANT----PPDQLLT 2875
Db 3363 SGLGRRTRALRVSHAFHSAHMDGMLDGFRAVLETVEFRRPSLTVVSNVTGLAAGPDDLCD 3422

```
Qy       2876 PHYWTQQARNTVDYATTTQTLHQHGVTTYIELGPDNTLTTLTHHNLPNPPTTTLT---LT 2932
Db       3423 PEYWVRHVRGTVRFLDGVRVLRDLGVRTCLELGPDGVLTAMAADGLADTPADSAAGSPVG 3482

Qy       2933 HPHHHPQTH---------LLTNL-----AKTTTTWHPHHYTH-HDNQPHTHT-------- 2969
Db       3483 SPAGSPADSAAGALRPRPLLVALLRRKRSETETVADALGRAHAHGTGPDWHAWFAGSGAH 3542

Qy       2970 HLDLPTYPFQHHHYWLESTQPGAG---NVSAAGLDPTEHPLLGATLELATDGGALLAGRL 3026
Db       3543 RVDLPTYSFRRDRYWLDA--PAADTAVDTAGLGLGTADHPLLGAVVSLPDRDGLLLTGRL 3600

Qy       3027 SLRSHPWLADHAVGGTVLLSGATFLELALHAGTYVGCDRVDELTLHAPLVVPVDGGVSVQ 3086
Db       3601 SLRTHPWLADHAVLGSVLLPGAAMVELAAHAAESAGLRDVRELTLLEPLVLPEHGGVELR 3660

Qy       3087 VGVAAADGE--------GRRLVSVYARGGSACGGGGASGGVWTCHASGVLVEAAAGGVVV 3138
Db       3661 VTVGAPAGEPGGESAGDGARPVSLHSRLADA-----PAGTAWSCHATGLLATDRPELPVA 3715

Qy       3139 DGLAGVWPPRGAVAVDVDGVRDRLAGAGCVLGPVFSGLRAVWRDGGDLLAEVCLP----- 3193
Db       3716 PDRAAMWPPQGAEEVPLDGLYERLDGNGLAFGPLFQGLNAVWRYEGEVFADIALPATTNA 3775

Qy       3194 -----EEAWGDAAG--FGLHPALLDGVVQPLSVLLPGGTGFGEGAGFGEGVRVPAVWGGV 3246
Db       3776 TAPATANGGGSAAAAPYGIHPALLDASLHAIAV---GGL-----VDEPELVRVPFHWSGV 3827

Qy       3247 SLHRAGVTGVRVRVSAVGRGGGREAVSVVVGDEAGVPVASVDRLELRPVDMGQLRAVSVS 3306
Db       3828 TVHAAGAAAARVRLAS----AGTDAVSLSLTDGEGRPLVSVERLTLRPVTADQAAASRVG 3883

Qy       3307 AGRRGSLYAVQW---------------AEVGPVPVCGQ--------------------A 3330
Db       3884 ----GLMHRVAWRPYALASSGEQDPHATSYGPTAVLGKDELKVAAALESAGVEVGLYPDL 3939

Qy       3331 WAWHEDVGESGGGPVPGVVVLRCPDAGAGGGGGGGGGGVGEVVGGVLGVVQGWLGLERF 3390
Db       3940 AALSQDV--AAGAPAPRTVLAPLP-----AGPADGGAEGVRGTVARTLELLQAWLADEHL 3992

Qy       3391 AGSRLVVVTRGAV----VAGPEDGPVDVVGASVWGLVRSAQAEHPDRFVLLDLDTDTGT- 3445
Db       3993 AGTRLLLVTRGAVRDPEGSGADDGGEDLSHAAAWGLVRTAQTENPGRFGLLDLADDASSY 4052

Qy       3446 ----DLDTGAGAGWGVDGGRVAAVVACGEPQLAVRGERLLAARLKRLESSGDVPAQRSGD 3501
Db       4053 RTLPSVLSDAGLR--------------DEPQLALHDGTIRLARL---------------- 4082

Qy       3502 TRARRSDVPAQRSGGVPARRSVDVSGREVLPWLSGGSVLVTGGTGVLGAAVARHLAGVCG 3561
Db       4083 -----ASVRPETGTAAPA----------LAP---EGTVLLTGGTGGLGGLVARHVVGEWG 4124

Qy       3562 VRDLLLVSRRGPDAPGAEGLRAELAALGAEVRIVACDVGERREVVRLLEGVPAGCPLTGV 3621
Db       4125 VRRLLLVSRRGTDAPGADELVHELEALGADVSVAACDVADREALTAVLDAIPAEHPLTAV 4184

Qy       3622 VHAAGVLDDATIASLTPERLGTVFAAKVDAALLLDELTR--GMELSAFVLFSSAAGILGS 3679
Db       4185 VHTAGVLSDGTLPSMTTEDVEHVLRPKVDAAFLLDELTSTPAYDLAAFVMFSSAAAVFGG 4244

Qy       3680 AGQGNYAAANAALDALAYRRRAAGLPGVSLAWGLWEEASGMTGHLAGTDHRRIIRSGLHP 3739
Db       4245 AGQGAYAAANATLDALAWRRRAAGLPALSLGWGLWAETSGMTGELGQADLRRMSRAGIGG 4304

Qy       3740 MSTPDALALFDAALALDR-PVLLP-----ADLRPA-----PPLPPLLQDLLPATRRR--- 3785
Db       4305 ISDAEGIALLDAALRDDRHPVLLPLRLDAAGLRDAAGNDPAGIPALFRDVVGARTVRARP 4364

Qy       3786 ------TTRTT--TTGGADNGAQLHA-----RLAGQTHEQQHTTLLALVRSHIATVLGHT 3832
Db       4365 SAASASTTAGTAGTPGTADGAAETAAVTLADRAATVDGPARQRLLLEFVVGEVAEVLGHA 4424

Qy       3833 TPDTIPPDRAFRDLGFDSLTAVELRNRLSRTTGLRLPTTLAFDHPNPTTLTHHLHTQL-- 3890
Db       4425 RGHRIDAERGFLDLGFDSLTAVELRNRLNSAGGLALPATLVFDHPSPAALASHLDAELPR 4484

Qy       3891 -QPQPDNA-----------------VAPVLAELDKLESALSALDKTDS-ASERVTLRLK 3930
Db       4485 GASDQDGAGNRNGNENGTTASRSTAETDALLAQLTRLEGALVLTGLSDAPGSEEVLEHLR 4544

Qy       3931 SLMLR------------------------WNAPQHPTAESADDD--EKFTSATEAEIFK 3963
Db       4545 SLRSMVTGETGTGTASGAPDGAGSGAEDRPWAAGDGAGGGSEDGAGVPDFMNASAEELFG 4604

Qy       3964 FIDND 3968
Db       4605 LLDQD 4609
```

RESULT 6
US-09-860-846-6
; Sequence 6, Application US/09860846
; Patent No. US20020164742A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/860,846
; CURRENT FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 6
; LENGTH: 11877
; TYPE: PRT
; ORGANISM: Streptomyces venezuelae
US-09-860-846-6

Query Match 39.1%; Score 8148; DB 9; Length 11877;
Best Local Similarity 44.8%; Pred. No. 0;
Matches 1909; Conservative 463; Mismatches 1307; Indels 586; Gaps 94;

```
Qy         15 GEVLGVADEADGGVVFVFPGQGPQWPGMGRELLDASDVFRESVRACEAAFAPYVDWSVEQ 74
Db        956 GLVRGVASGV-GRVAFVFPGQGTQWAGMGAELLDSSAVFAAAMAECEAALSPYVDWSLEA 1014

Qy         75 VLRDSPDAPGLDRVDVVQPTLFAVMISLAALWRSQGVEPCAVLGHSLGEIAAAHVSGGLS 134
Db        1015 VVRQAPGAPTLERVDVVQPVTFAVMVSLARVWQHHGVTPQAVVGHSQGEIAAAYVAGALS 1074

Qy        135 LADAARVVTLWSQA-QTTLAGTGALVSVAATPDELLPRIAPWTEDNPARLAVAAVNGPRS 193
Db        1075 LDDAARVVTLRSKSIAAHLAGKGGMLSLALSEDAVLERLAGFD-----GLSVAAVNGPTA 1129

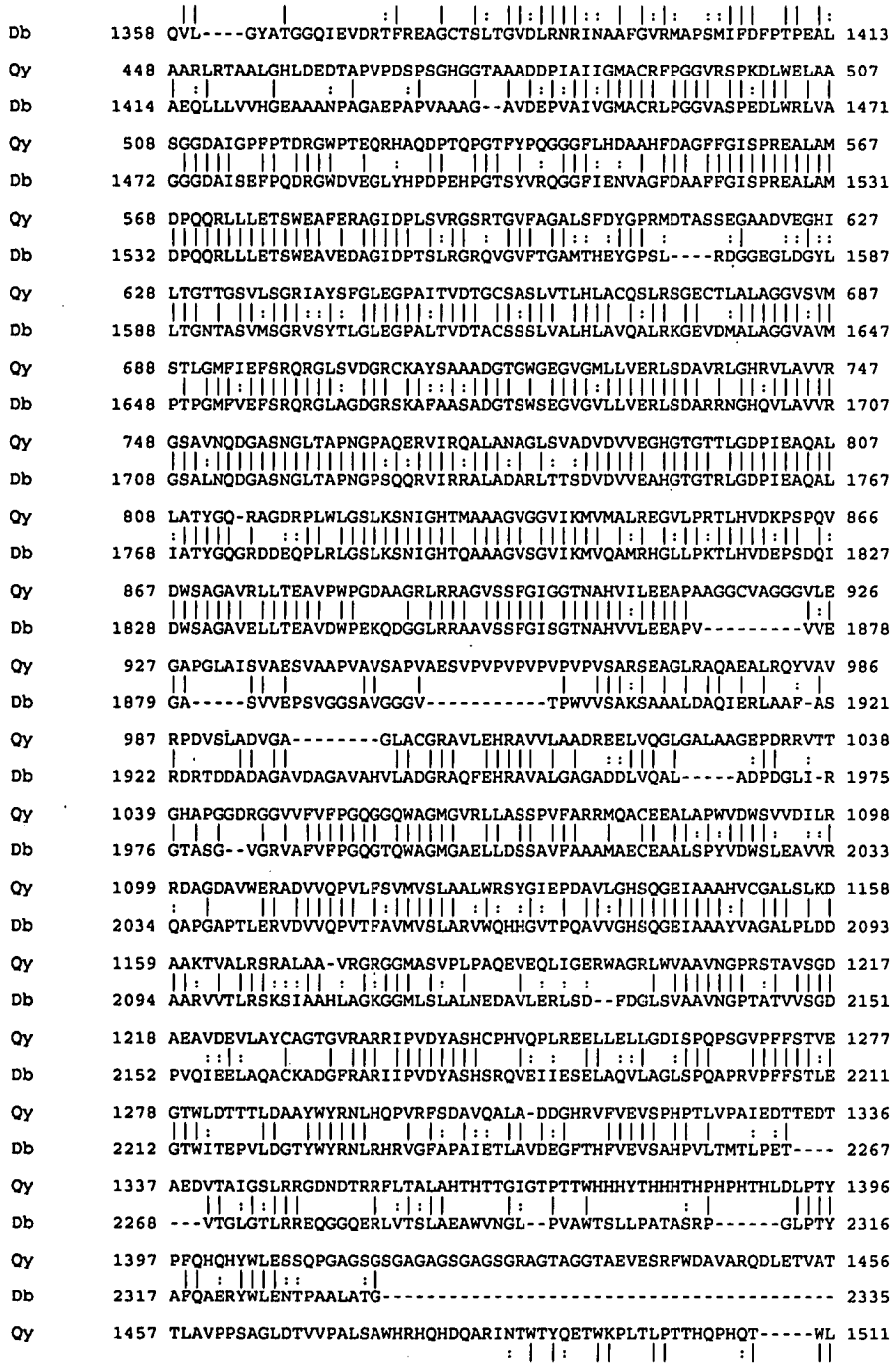
Qy        194 TVVSGAREAVADLVADLTAAQVRTRMIPVDVPAHSPLMYAIEERVVSGLLPITPRPSRIP 253
Db        1130 TVVSGDPVQIEELARACEADGVRARVIPVDYASHSRQVEIIESELAEVLAGLSPQAPRVP 1189

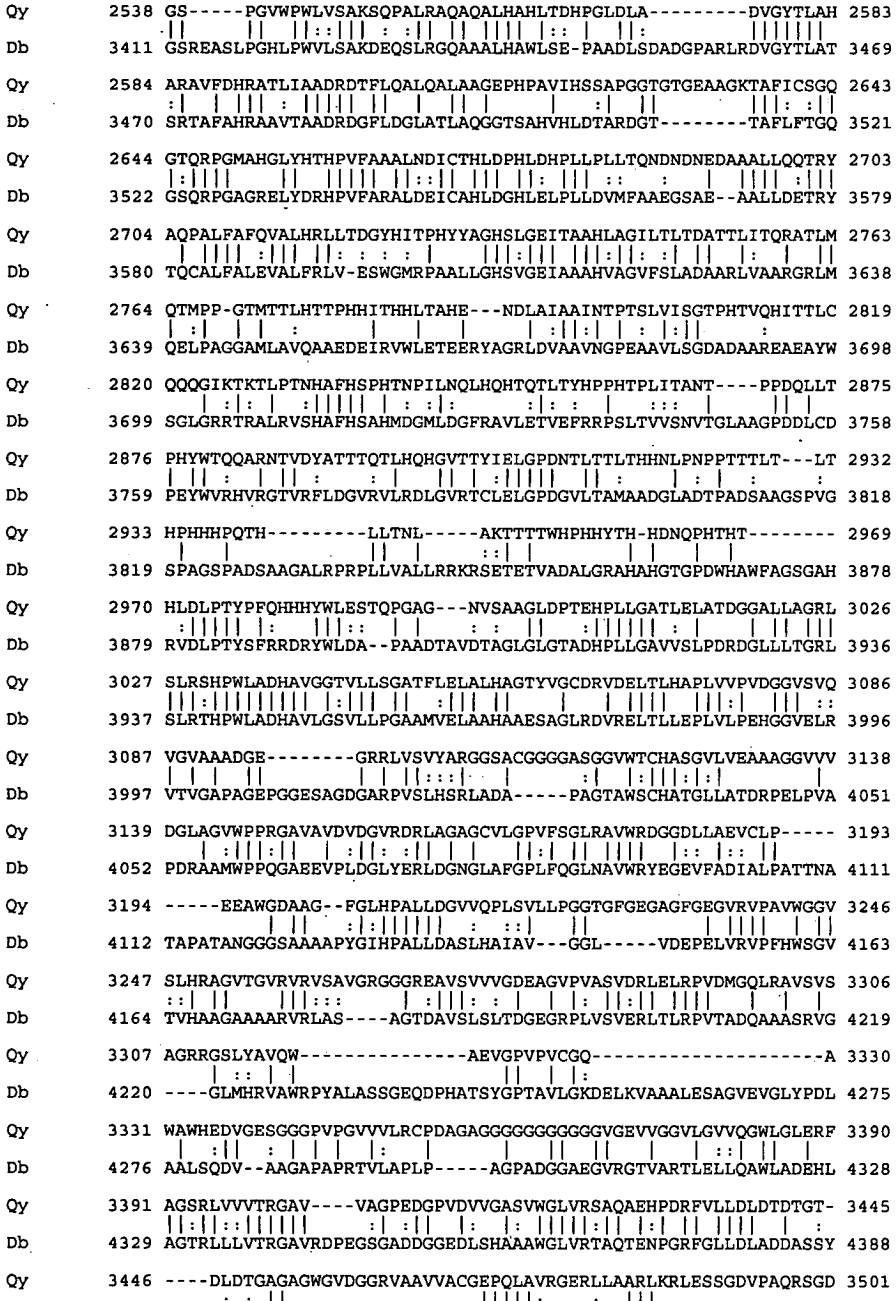
Qy        254 FHSSVTGGRLDTRELDAAYWYRNMSSTVRFEPAAR-LLLQQGPKTFVEMSPHPVLTMGLQ 312
Db        1190 FFSTLEGAWITEPVLDGGYWYRNLRHRVGFAPAVETLATDEGFTHFVEVSAHPVLTMALP 1249

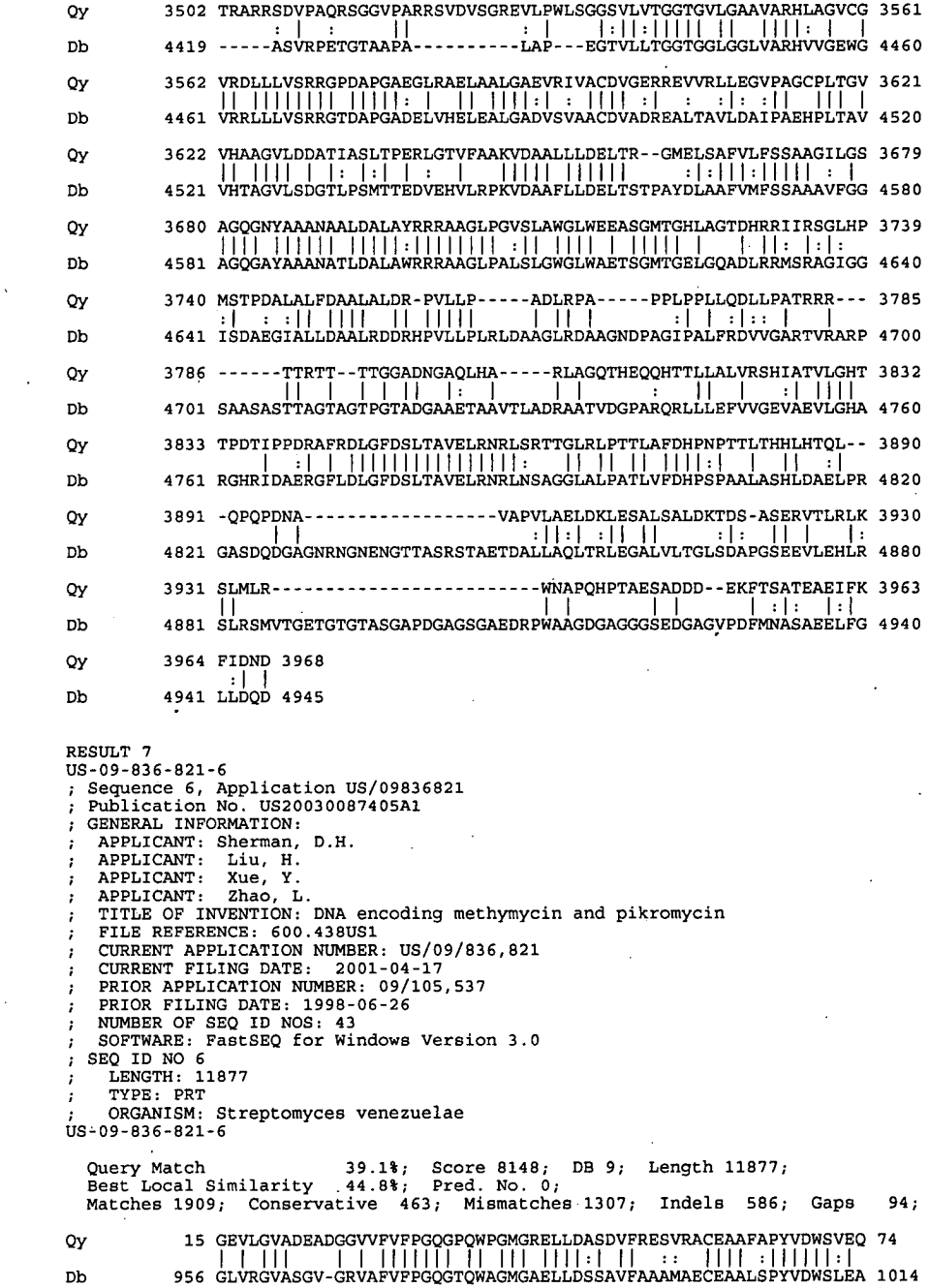
Qy        313 ELAPDLGDTTGTADTVIMGTLRRGQGTLDHFLTSLAQ----------------LRGHGET 356
Db        1250 ------GTVTGLA------TLRRDNGGQDRLVASLAEAWANGLAVDWSPLLPSATGHHSD 1297

Qy        357 SATTVLS------ARLTALSPT----------------------QQQSLLLDLVRAHTM 387
Db        1298 LPTYAFQTERHWLGEIEALAPAGEPAVQPAVLRTEAAEPAELDRDEQLRVILDKVRAQTA 1357

Qy        388 AVLNDDGNERTASDAGPSASFAHLGFDSVMGVELRNRLSKATGLRLPVTLIFDHTTPAAV 447
```

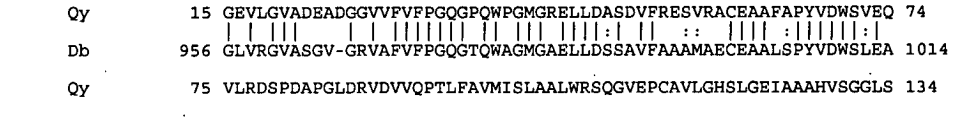

Db 1358 QVL----GYATGGQIEVDRTFREAGCTSLTGVDLRNRINAAFGVRMAPSMIFDFPTPEAL 1413
Qy 448 AARLRTAALGHLDEDTAPVPDSPSGHGGTAAADDPIAIIGMACRFPGGVRSPKDLWELAA 507
Db 1414 AEQLLLVVHGEAAANPAGAEPAPVAAAG--AVDEPVAIVGMACRLPGGVASPEDLWRLVA 1471
Qy 508 SGGDAIGPFPTDRGWPTEQRHAQDPTQPGTFYPQGGGFLHDAAHFDAGFFGISPREALAM 567
Db 1472 GGGDAISEFPQDRGWDVEGLYHPDPEHPGTSYVRQGGFIENVAGFDAAFFGISPREALAM 1531
Qy 568 DPQQRLLLETSWEAFERAGIDPLSVRGSRTGVFAGALSFDYGPRMDTASSEGAADVEGHI 627
Db 1532 DPQQRLLLETSWEAVEDAGIDPTSLRGRQVGVFTGAMTHEYGPSL----RDGGEGLDGYL 1587
Qy 628 LTGTTGSVLSGRIAYSFGLEGPAITVDTGCSASLVTLHLACQSLRSGECTLALAGGVSVM 687
Db 1588 LTGNTASVMSGRVSYTLGLEGPALTVDTACSSSLVALHLAVQALRKGEVDMALAGGVAVM 1647
Qy 688 STLGMFIEFSRQRGLSVDGRCKAYSAAADGTGWGEGVGMLLVERLSDAVRLGHRVLAVVR 747
Db 1648 PTPGMFVEFSRQRGLAGDGRSKAFAASADGTSWSEGVGVLLVERLSDARRNGHQVLAVVR 1707
Qy 748 GSAVNQDGASNGLTAPNGPAQERVIRQALANAGLSVADVDVVEGHGTGTTLGDPIEAQAL 807
Db 1708 GSALNQDGASNGLTAPNGPSQQRVIRRALADARLTTSDVDVVEAHGTGTRLGDPIEAQAL 1767
Qy 808 LATYGQ-RAGDRPLWLGSLKSNIGHTMAAAGVGGVIKMVMALREGVLPRTLHVDKPSPQV 866
Db 1768 IATYGQGRDDEQPLRLGSLKSNIGHTQAAAGVSGVIKMVQAMRHGLLPKTLHVDEPSDQI 1827
Qy 867 DWSAGAVRLLTEAVPWPGDAAGRLRRAGVSSFGIGGTNAHVILEEAPAAGGCVAGGGVLE 926
Db 1828 DWSAGAVELLTEAVDWPEKQDGGLRRAAVSSFGISGTNAHVVLEEAPV---------VVE 1878
Qy 927 GAPGLAISVAESVAAPVAVSAPVAESVPVPVPVPVPVPVSARSEAGLRAQAEALRQYVAV 986
Db 1879 GA-----SVVEPSVGGSAVGGGV-----------TPWVVSAKSAAALDAQIERLAAF-AS 1921
Qy 987 RPDVSLADVGA--------GLACGRAVLEHRAVVLAADREELVQGLGALAAGEPDRRVTT 1038
Db 1922 RDRTDDADAGAVDAGAVAHVLADGRAQFEHRAVALGAGADDLVQAL-----ADPDGLI-R 1975
Qy 1039 GHAPGGDRGGVVFVFPGQGGQWAGMGVRLLASSPVFARRMQACEEALAPWVDWSVVDILR 1098
Db 1976 GTASG--VGRVAFVFPGQGTQWAGMGAELLDSSAVFAAAMAECEAALSPYVDWSLEAVVR 2033
Qy 1099 RDAGDAVWERADVVQPVLFSVMVSLAALWRSYGIEPDAVLGHSQGEIAAAHVCGALSLKD 1158
Db 2034 QAPGAPTLERVDVVQPVTFAVMVSLARVWQHHGVTPQAVVGHSQGEIAAAYVAGALPLDD 2093
Qy 1159 AAKTVALRSRALAA-VRGRGGMASVPLPAQEVEQLIGERWAGRLWVAAVNGPRSTAVSGD 1217
Db 2094 AARVVTLRSKSIAAHLAGKGGMLSLALNEDAVLERLSD--FDGLSVAAVNGPTATVVSGD 2151
Qy 1218 AEAVDEVLAYCAGTGVRARRIPVDYASHCPHVQPLREELLELLGDISPQPSGVPFFSTVE 1277
Db 2152 PVQIEELAQACKADGFRARIIPVDYASHSRQVEIIESELAQVLAGLSPQAPRVPFFSTLE 2211
Qy 1278 GTWLDTTTLDAAYWYRNLHQPVRFSDAVQALA-DDGHRVFVEVSPHPTLVPAIEDTTEDT 1336
Db 2212 GTWITEPVLDGTYWYRNLRHRVGFAPAIETLAVDEGFTHFVEVSAHPVLTMTLPET---- 2267
Qy 1337 AEDVTAIGSLRRGDNDTRRFLTALAHTHTTGIGTPTTWHHHYTHHHTHPHPHTHLDLPTY 1396
Db 2268 ---VTGLGTLRREQGGQERLVTSLAEAWVNGL--PVAWTSLLPATASRP------GLPTY 2316
Qy 1397 PFQHQHYWLESSQPGAGSGSGAGAGSGAGSGRAGTAGGTAEVESRFWDAVARQDLETVAT 1456
Db 2317 AFQAERYWLENTPAALATG---------- 2335
Qy 1457 TLAVPPSAGLDTVVPALSAWHRHQHDQARINTWTYQETWKPLTLPTTHQPHQT-----WL 1511
Db 2336 ----------DDWRYRIDWK--RLPAAEGSERTGLSGRWL 2363
Qy 1512 IAIPETQTHHPHITNILTNLHHHGITPIPLTLNHTHTNPQHLHHTLHHTRQQAQNHTTG- 1570
Db 2364 AVTPE--DHSAQAAAVLTALVDAGAKVEVLTAG-ADDDREALAARLTAL-------TTGD 2413
Qy 1571 AITGLLSLLALDETPHPHHPHTPTGTLLNLTLTQTHTQTHPPTPLWYATTNATTTHPNDP 1630
Db 2414 GFTGVVSLL----------DGLVPQVAWVQALGDAGIKAPLWSVTQGAVSVGRLDT 2459
Qy 1631 LTHPTQAQTWGLARTTLLEHPTHTAGIIDLPTTPTPHTLQHLTQTLTQPHHQTQLAIRTT 1690
Db 2460 PADPDRAMLWGLGRVVALEHPERWAGLVDLPAQPDAAALAHLVTALSGATGEDQIAIRTT 2519
Qy 1691 GTHTRRLTPTTLTPTH-QPPTP--TPHGTTLITGGTGALATHLTHHLTTHQPTQHLLLTS 1747
Db 2520 GLHARRL---ARAPLHGRRPTRDWQPHGTVLITGGTGALGSHAARWM-AHHGAEHLLLVS 2575
Qy 1748 RTGPHTPHAQHLTTQLQQKGIHLTITTCDTSNPDQLQQLLNTIPPQHPLTTVIHTAGILD 1807
Db 2576 RSGEQAPGATQLTAELTASGARVTIAACDVADPHAMRTLLDAIPAETPLTAVVHTAGALD 2635
Qy 1808 DATLTNLTPTQLNNVLRAKAHSAHLLHQLTQHTPLTAFVLYSSAAATFGAPGQANYAAAN 1867
Db 2636 DGIVDTLTAEQVRRAHRAKAVGASVLDELTRDLDLDAFVLFSSVSSTLGIPGQGNYAPHN 2695
Qy 1868 AYLDALAHHRHTHHLPATSIAWGTWQGNGLADSDKARAYLDRRGFRPMSPELATAAVTQA 1927
Db 2696 AYLDALAARRRATGRSAVSVAWGPWDGGGMAAGDGVAERLRNHGVPGMDPELALAALESA 2755
Qy 1928 IADTERPYVVIADIDWSK-------------IEHTSQTSDLVSA-----ARE-REPAVQRP 1969
Db 2756 LGRDETA-ITVADIDWDRFYLAYSSGRPQPLVEELPEVRRIIDARDSATSGQGGSSAQGA 2814
Qy 1970 TPPAELHKTLAHQTSADQRAALLELVRDHVAAVLRHADPKAIAPDQSFRALGFDSLTAVE 2029
Db 2815 NPLAE---RLAAAAPGERTEILLGLVRAQAAAVLRMRSPEDVAADRAFKDIGFDSLAGVE 2871
Qy 2030 FRNLLIKATGLRLPVSLVFDHPTPAKLAVHLQNQLRG-----TAAESAPSAAAVTAEASV 2084
Db 2872 LRNRLTRATGLQLPATLVFDHPTPLALVSLLRSEFLGDEETADARRSAALPATVGAGAGA 2931
Qy 2085 -------TEPIAIVGMACRFPGGVTSADDFWDLISSEQDAIGGFPTDRGWDLDTLYDPDP 2137
Db 2932 GAGTDADDDPIAIVAMSCRYPGDIRSPEDLWRMLSEGGEGITPFPTDRGWDLDGLYDADP 2991
Qy 2138 DHPGTCYTRNGGFLYDAGHFDAEFFGISPREALAMDPQQRLLLETAWETIEHAGINPHTL 2197
Db 2992 DALGRAYVREGGFLHDAAEFDAEFFGVSPREALAMDPQQRMLLTTSWEAFERAGIEPASL 3051
Qy 2198 HGTPTGVFTGTNGQDYALRVHNAGQSTDGFALTGTAGSVISGRISYTFGFEGPAVSVDTA 2257
Db 3052 RGSSTGVFIGLSYQDYAARVPNAPRGVEGYLLTGSTPSVASGRIAYTFGLEGPATTVDTA 3111
Qy 2258 CSSSLVALHLACQALRAGECSMALAGGVTVMSSPGAFVEFSRQRGLAADGHCKAFSAAAD 2317
Db 3112 CSSSLTALHLAVRALRSGECTMALAGGVAMMATPHMFVEFSRQRALAPDGRSKAFSADAD 3171
Qy 2318 GTGWGEGVGMLLVERLSDAHRNGHRVLAVVRGSAVNQDGASNGLTAPNGPSQQRVIRQAL 2377
Db 3172 GFGAAEGVGLLLVERLSDARRNGHPVLAVVRGTAVNQDGASNGLTAPNGPSQQRVIRQAL 3231
Qy 2378 ANAGLSAGDVDAVEAHGTGTTLGDPIEAQALLATYGQDRAGEGPLWLGSVKSNVGHTQAA 2437
Db 3232 ADARLAPGDIDAVETHGTGTSLGDPIEAQGLQATYGKERPAERPLAIGSVKSNIGHTQAA 3291
Qy 2438 AGVAGVIKMVMALRHGLLPRTLHVDEPSPHVDWSAGAVQLLTETVPWPGGEGRLRRAGVS 2497
Db 3292 AGAAGIIKMVLAMRHGTLPKTLHADEPSPHVDWANSGLALVTEPIDWPAGTGP-RRAAVS 3350
Qy 2498 SFGVSGTNAHVILEEAP--------ADDVP-----------GGPPAGEGDAGSDDEAAA 2537
Db 3351 SFGISGTNAHVVLEQAPDAAGEVLGADEVPEVSETVAMAGTAGTSEVAEGSEASEAPAAP 3410

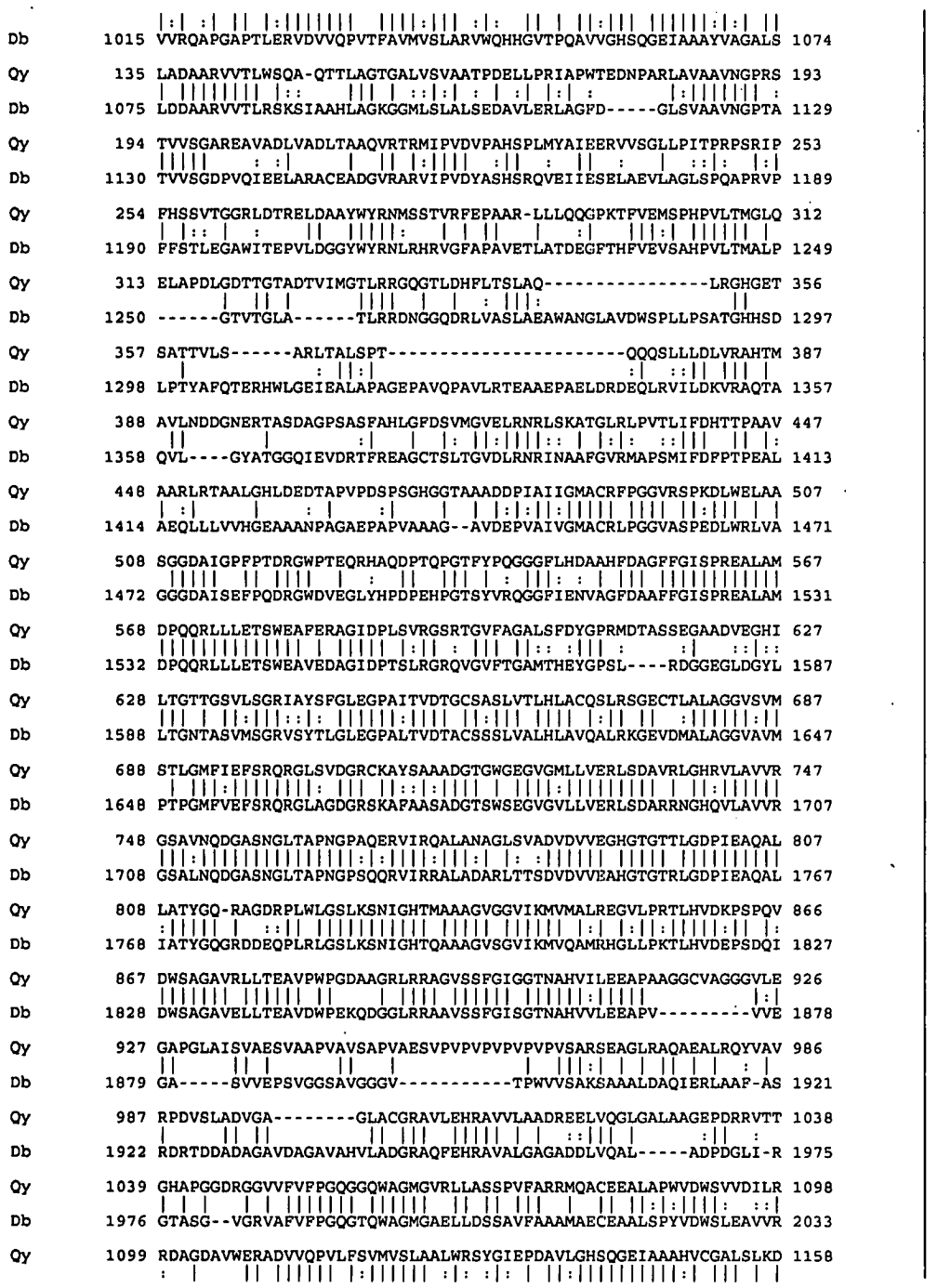


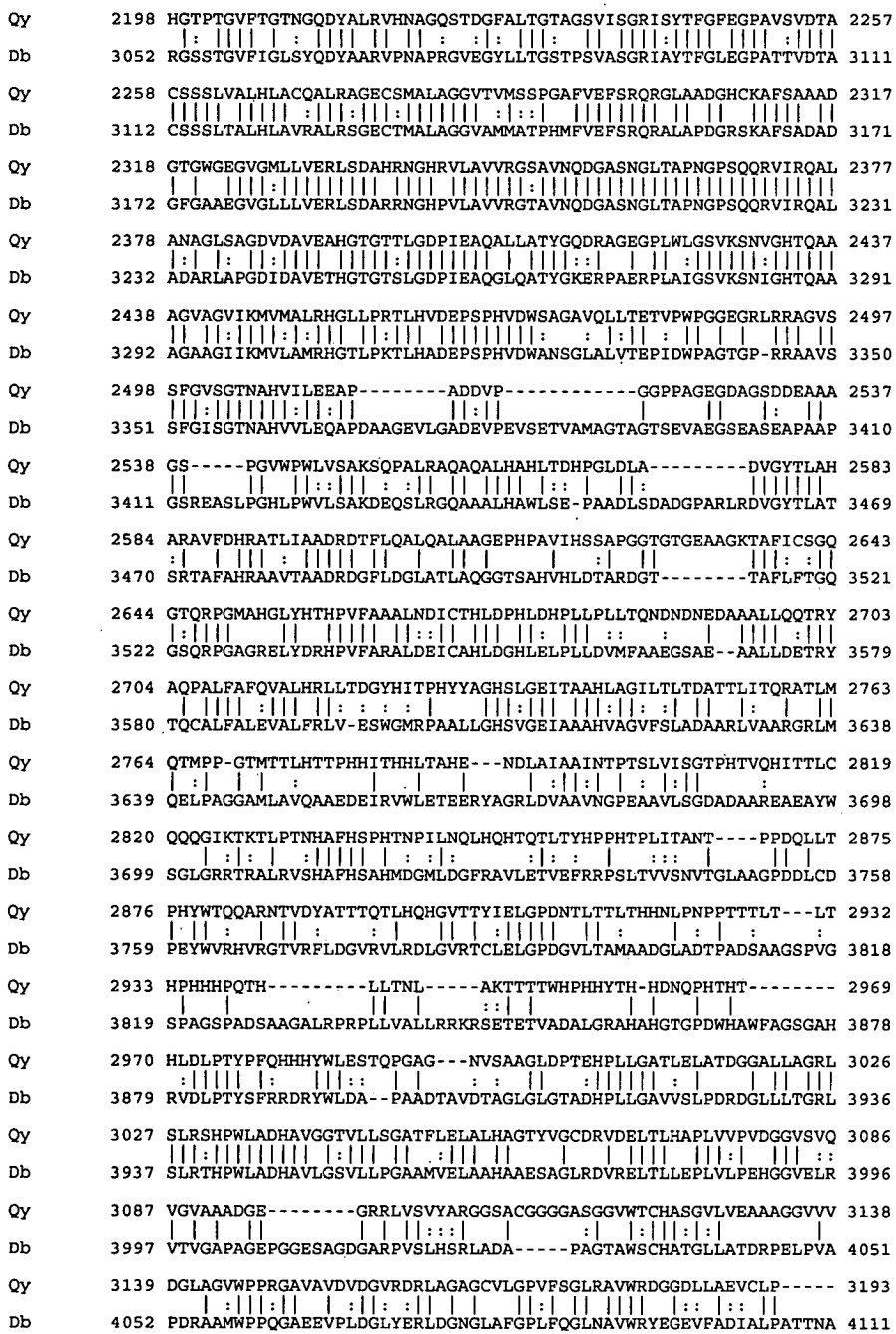


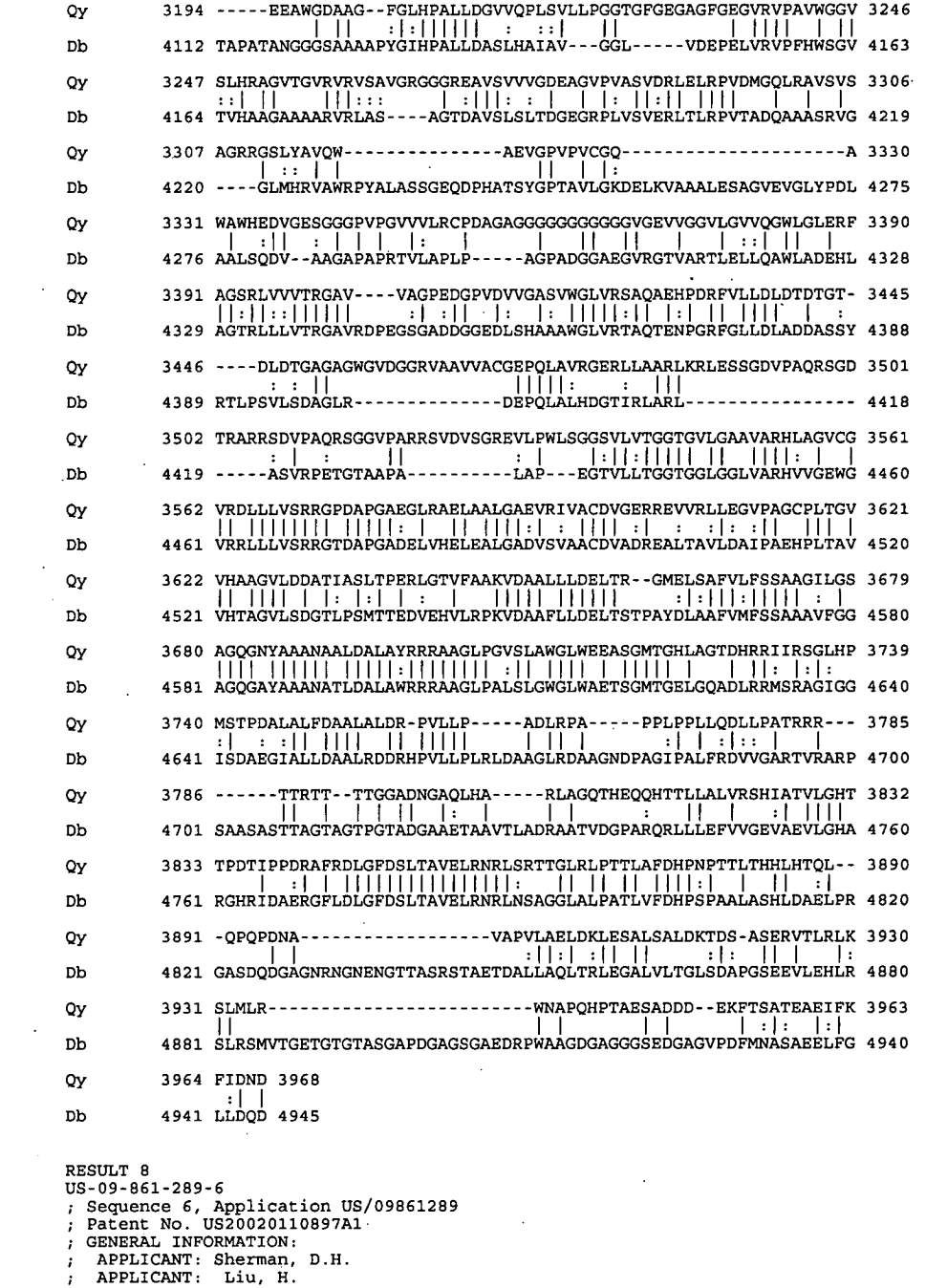
RESULT 7
US-09-836-821-6
; Sequence 6, Application US/09836821
; Publication No. US20030087405A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/836,821
; CURRENT FILING DATE: 2001-04-17
; PRIOR APPLICATION NUMBER: 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 6
; LENGTH: 11877
; TYPE: PRT
; ORGANISM: Streptomyces venezuelae
US-09-836-821-6

Query Match 39.1%; Score 8148; DB 9; Length 11877;
Best Local Similarity 44.8%; Pred. No. 0;
Matches 1909; Conservative 463; Mismatches 1307; Indels 586; Gaps 94;




Db 1015 VVRQAPGAPTLERVDVVQPVTFAVMVSLARVWQHHGVTPQAVVGHSQGEIAAAYVAGALS 1074
Qy 135 LADAARVVTLWSQA-QTTLAGTGALVSVAATPDELLPRIAPWTEDNPARLAVAAVNGPRS 193
Db 1075 LDDAARVVTLRSKSIAAHLAGKGGMLSLALSEDAVLERLAGFD-----GLSVAAVNGPTA 1129
Qy 194 TVVSGAREAVADLVADLTAAQVRTRMIPVDVPAHSPLMYAIEERVVSGLLPITPRPSRIP 253
Db 1130 TVVSGDPVQIEELARACEADGVRARVIPVDYASHSRQVEIIESELAEVLAGLSPQAPRVP 1189
Qy 254 FHSSVTGGRLDTRELDAAYWYRNMSSTVRFEPAAR-LLLQQGPKTFVEMSPHPVLTMGLQ 312
Db 1190 FFSTLEGAWITEPVLDGGYWYRNLRHRVGFAPAVETLATDEGFTHFVEVSAHPVLTMALP 1249
Qy 313 ELAPDLGDTTGTADTVIMGTLRRGQGTLDHPLTSLAQ----------------LRGHGET 356
Db 1250 ------GTVTGLA------TLRRDNGGQDRLVASLAEAWANGLAVDWSPLLPSATGHHSD 1297
Qy 357 SATTVLS------ARLTALSPT----------------------QQQSLLLDLVRAHTM 387
Db 1298 LPTYAFQTERHWLGEIEALAPAGEPAVQPAVLRTEAAEPAELDRDEQLRVILDKVRAQTA 1357
Qy 388 AVLNDDGNERTASDAGPSASFAHLGFDSVMGVELRNRLSKATGLRLPVTLIFDHTTPAAV 447
Db 1358 QVL----GYATGGQIEVDRTFREAGCTSLTGVDLRNRINAAFGVRMAPSMIFDFPTPEAL 1413
Qy 448 AARLRTAALGHLDEDTAPVPDSPSGHGGTAAADDPIAIIGMACRFPGGVRSPKDLWELAA 507
Db 1414 AEQLLLVVHGEAAANPAGAEPAPVAAAG--AVDEPVAIVGMACRLPGGVASPEDLWRLVA 1471
Qy 508 SGGDAIGPFPTDRGWPTEQRHAQDPTQPGTFYPQGGGFLHDAAHFDAGFFGISPREALAM 567
Db 1472 GGGDAISEFPQDRGWDVEGLYHPDPEHPGTSYVRQGGFIENVAGFDAAFFGISPREALAM 1531
Qy 568 DPQQRLLLETSWEAFERAGIDPLSVRGSRTGVFAGALSFDYGPRMDTASSEGAADVEGHI 627
Db 1532 DPQQRLLLETSWEAVEDAGIDPTSLRGRQVGVFTGAMTHEYGPSL----RDGGEGLDGYL 1587
Qy 628 LTGTTGSVLSGRIAYSFGLEGPAITVDTGCSASLVTLHLACQSLRSGECTLALAGGVSVM 687
Db 1588 LTGNTASVMSGRVSYTLGLEGPALTVDTACSSSLVALHLAVQALRKGEVDMALAGGVAVM 1647
Qy 688 STLGMFIEFSRQRGLSVDGRCKAYSAAADGTGWGEGVGMLLVERLSDAVRLGHRVLAVVR 747
Db 1648 PTPGMFVEFSRQRGLAGDGRSKAFAASADGTSWSEGVGVLLVERLSDARRNGHQVLAVVR 1707
Qy 748 GSAVNQDGASNGLTAPNGPAQERVIRQALANAGLSVADVDVVEGHGTGTTLGDPIEAQAL 807
Db 1708 GSALNQDGASNGLTAPNGPSQQRVIRRALADARLTTSDVDVVEAHGTGTRLGDPIEAQAL 1767
Qy 808 LATYGQ-RAGDRPLWLGSLKSNIGHTMAAAGVGGVIKMVMALREGVLPRTLHVDKPSPQV 866
Db 1768 IATYGQGRDDEQPLRLGSLKSNIGHTQAAAGVSGVIKMVQAMRHGLLPKTLHVDEPSDQI 1827
Qy 867 DWSAGAVRLLTEAVPWPGDAAGRLRRAGVSSFGIGGTNAHVILEEAPAAGGCVAGGGVLE 926
Db 1828 DWSAGAVELLTEAVDWPEKQDGGLRRAAVSSFGISGTNAHVVLEEAPV---------VVE 1878
Qy 927 GAPGLAISVAESVAAPVAVSAPVAESVPVPVPVPVPVPVSARSEAGLRAQAEALRQYVAV 986
Db 1879 GA-----SVVEPSVGGSAVGGGV-----------TPWVVSAKSAAALDAQIERLAAF-AS 1921
Qy 987 RPDVSLADVGA--------GLACGRAVLEHRAVVLAADREELVQGLGALAAGEPDRRVTT 1038
Db 1922 RDRTDDADAGAVDAGAVAHVLADGRAQFEHRAVALGAGADDLVQAL-----ADPDGLI-R 1975
Qy 1039 GHAPGGDRGGVVFVFPGQGGQWAGMGVRLLASSPVFARRMQACEEALAPWVDWSVVDILR 1098
Db 1976 GTASG--VGRVAFVFPGQGTQWAGMGAELLDSSAVFAAAMAECEAALSPYVDWSLEAVVR 2033
Qy 1099 RDAGDAVWERADVVQPVLFSVMVSLAALWRSYGIEPDAVLGHSQGEIAAAHVCGALSLKD 1158
Db 2034 QAPGAPTLERVDVVQPVTFAVMVSLARVWQHHGVTPQAVVGHSQGEIAAAYVAGALPLDD 2093
Qy 1159 AAKTVALRSRALAA-VRGRGGMASVPLPAQEVEQLIGERWAGRLWVAAVNGPRSTAVSGD 1217
Db 2094 AARVVTLRSKSIAAHLAGKGGMLSLALNEDAVLERLSD--FDGLSVAAVNGPTATVVSGD 2151
Qy 1218 AEAVDEVLAYCAGTGVRARRIPVDYASHCPHVQPLREELLELLGDISPQPSGVPFFSTVE 1277
Db 2152 PVQIEELAQACKADGFRARIIPVDYASHSRQVEIIESELAQVLAGLSPQAPRVPFFSTLE 2211
Qy 1278 GTWLDTTTLDAAYWYRNLHQPVRFSDAVQALA-DDGHRVFVEVSPHPTLVPAIEDTTEDT 1336
Db 2212 GTWITEPVLDGTYWYRNLRHRVGFAPAIETLAVDEGFTHFVEVSAHPVLTMTLPET---- 2267
Qy 1337 AEDVTAIGSLRRGDNDTRRFLTALAHTHTTGIGTPTTWHHHYTHHHTHPHPHTHLDLPTY 1396
Db 2268 ---VTGLGTLRREQGGQERLVTSLAEAWVNGL--PVAWTSLLPATASRP------GLPTY 2316
Qy 1397 PFQHQHYWLESSQPGAGSGSGAGAGSGAGSGRAGTAGGTAEVESRFWDAVARQDLETVAT 1456
Db 2317 AFQAERYWLENTPAALATG-- 2335
Qy 1457 TLAVPPSAGLDTVVPALSAWHRHQHDQARINTWTYQETWKPLTLPTTHQPHQT-----WL 1511
Db 2336 ----------------------------DDWRYRIDWK--RLPAAEGSERTGLSGRWL 2363
Qy 1512 IAIPETQTHHPHITNILTNLHHHGITPIPLTLNHTHTNPQHLHHTLHHTRQQAQNHTTG- 1570
Db 2364 AVTPE--DHSAQAAAVLTALVDAGAKVEVLTAG-ADDDREALAARLTAL-------TTGD 2413
Qy 1571 AITGLLSLLALDETPHPHHPHTPTGTLLNLTLTQTHTQTHPPTPLWYATTNATTTHPNDP 1630
Db 2414 GFTGVVSLL--------------DGLVPQVAWVQALGDAGIKAPLWSVTQGAVSVGRLDT 2459
Qy 1631 LTHPTQAQTWGLARTTLLEHPTHTAGIIDLPTTPTPHTLQHLTQTLTQPHHQTQLAIRTT 1690
Db 2460 PADPDRAMLWGLGRVVALEHPERWAGLVDLPAQPDAAALAHLVTALSGATGEDQIAIRTT 2519
Qy 1691 GTHTRRLTPTTLTPTH-QPPTP--TPHGTTLITGGTGALATHLTHHLTTHQPTQHLLLTS 1747
Db 2520 GLHARRL---ARAPLHGRRPTRDWQPHGTVLITGGTGALGSHAARWM-AHHGAEHLLLVS 2575
Qy 1748 RTGPHTPHAQHLTTQLQQKGIHLTITTCDTSNPDQLQQLLNTIPPQHPLTTVIHTAGILD 1807
Db 2576 RSGEQAPGATQLTAELTASGARVTIAACDVADPHAMRTLLDAIPAETPLTAVVHTAGALD 2635
Qy 1808 DATLTNLTPTQLNNVLRAKAHSAHLLHQLTQHTPLTAFVLYSSAAATFGAPGQANYAAAN 1867
Db 2636 DGIVDTLTAEQVRRAHRAKAVGASVLDELTRDLDLDAFVLFSSVSSTLGIPGQGNYAPHN 2695
Qy 1868 AYLDALAHHRHTHHLPATSIAWGTWQGNGLADSDKARAYLDRRGFRPMSPELATAAVTQA 1927
Db 2696 AYLDALAARRRATGRSAVSVAWGPWDGGGMAAGDGVAERLRNHGVPGMDPELALAALESA 2755
Qy 1928 IADTERPYVVIADIDWSK-------------IEHTSQTSDLVSA-----ARELREPAVQRP 1969
Db 2756 LGRDETA-ITVADIDWDRFYLAYSSGRPQPLVEELPEVRRIIDARDSATSGQGGSSAQGA 2814
Qy 1970 TPPAELHKTLAHQTSADQRAALLELVRDHVAAVLRHADPKAIAPDQSFRALGFDSLTAVE 2029
Db 2815 NPLAE---RLAAAAPGERTEILLGLVRAQAAAVLRMRSPEDVAADRAFKDIGFDSLAGVE 2871
Qy 2030 FRNLLIKATGLRLPVSLVFDHPTPAKLAVHLQNQLRG-----TAAESAPSAAAVTAEASV 2084
Db 2872 LRNRLTRATGLQLPATLVFDHPTPLALVSLLRSEFLGDEETADARRSAALPATVGAGAGA 2931
Qy 2085 -------TEPIAIVGMACRFPGGVTSADDFWDLISSEQDAIGGFPTDRGWDLDTLYDPDP 2137
Db 2932 GAGTDADDDPIAIVAMSCRYPGDIRSPEDLWRMLSEGGEGITPFPTDRGWDLDGLYDADP 2991
Qy 2138 DHPGTCYTRNGGFLYDAGHFDAEFFGISPREALAMDPQQRLLLETAWETIEHAGINPHTL 2197
Db 2992 DALGRAYVREGGFLHDAAEFDAEFFGVSPREALAMDPQQRMLLTTSWEAFERAGIEPASL 3051





RESULT 8
US-09-861-289-6
; Sequence 6, Application US/09861289
; Patent No. US20020110897A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.

```
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/861,289
;  CURRENT FILING DATE:  2001-05-18
;  PRIOR APPLICATION NUMBER: 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 6
;   LENGTH: 11877
;   TYPE: PRT
;   ORGANISM: Streptomyces venezuelae
US-09-861-289-6

  Query Match             39.1%;  Score 8148;  DB 10;  Length 11877;
  Best Local Similarity   44.8%;  Pred. No. 0;
  Matches 1909;  Conservative  463;  Mismatches 1307;  Indels  586;  Gaps   94;

Qy         15 GEVLGVADEADGGVVFVFPGQGPQWPGMGRELLDASDVFRESVRACEAAFAPYVDWSVEQ  74
Db        956 GLVRGVASGV-GRVAFVFPGQGTQWAGMGAELLDSSAVFAAAMAECEAALSPYVDWSLEA 1014

Qy         75 VLRDSPDAPGLDRVDVVQPTLFAVMISLAALWRSQGVEPCAVLGHSLGEIAAAHVSGGLS 134
Db       1015 VVRQAPGAPTLERVDVVQPVTFAVMVSLARVWQHHGVTPQAVVGHSQGEIAAAYVAGALS 1074

Qy        135 LADAARVVTLWSQA-QTTLAGTGALVSVAATPDELLPRIAPWTEDNPARLAVAAVNGPRS 193
Db       1075 LDDAARVVTLRSKSIAAHLAGKGGMLSLALSEDAVLERLAGFD-----GLSVAAVNGPTA 1129

Qy        194 TVVSGAREAVADLVADLTAAQVRTRMIPVDVPAHSPLMYAIEERVVSGLLPITPRPSRIP 253
Db       1130 TVVSGDPVQIEELARACEADGVRARVIPVDYASHSRQVEIIESELAEVLAGLSPQAPRVP 1189

Qy        254 FHSSVTGGRLDTRELDAAYWYRNMSSTVRFEPAAR-LLLQQGPKTFVEMSPHPVLTMGLQ 312
Db       1190 FFSTLEGAWITEPVLDGGYWYRNLRHRVGFAPAVETLATDEGFTHFVEVSAHPVLTMALP 1249

Qy        313 ELAPDLGDTTGTADTVIMGTLRRGQGTLDHPLTSLAQ---------------LRGHGET 356
Db       1250 ------GTVTGLA------TLRRDNGGQDRLVASLAEAWANGLAVDWSPLLPSATGHHSD 1297

Qy        357 SATTVLS------ARLTALSPT----------------------QQQSLLLDLVRAHTM 387
Db       1298 LPTYAFQTERHWLGEIEALAPAGEPAVQPAVLRTEAAEPAELDRDEQLRVILDKVRAQTA 1357

Qy        388 AVLNDDGNERTASDAGPSASFAHLGFDSVMGVELRNRLSKATGLRLPVTLIFDHTTPAAV 447
Db       1358 QVL----GYATGGQIEVDRTFREAGCTSLTGVDLRNRINAAFGVRMAPSMIFDFPTPEAL 1413

Qy        448 AARLRTAALGHLDEDTAPVPDSPSGHGGTAAADDPIAIIGMACRFPGGVRSPKDLWELAA 507
Db       1414 AEQLLLVVHGEAAANPAGAEPAPVAAAG--AVDEPVAIVGMACRLPGGVASPEDLWRLVA 1471

Qy        508 SGGDAIGPFPTDRGWPTEQRHAQDPTQPGTFYPQGGGFLHDAAHFDAGFFGISPREALAM 567
Db       1472 GGGDAISEFPQDRGWDVEGLYHPDPEHPGTSYVRQGGFIENVAGFDAAFFGISPREALAM 1531

Qy        568 DPQQRLLLETSWEAFERAGIDPLSVRGSRTGVFAGALSFDYGPRMDTASSEGAADVEGHI 627
Db       1532 DPQQRLLLETSWEAVEDAGIDPTSLRGRQVGVFTGAMTHEYGPSL----RDGGEGLDGYL 1587

Qy        628 LTGTTGSVLSGRIAYSFGLEGPAITVDTGCSASLVTLHLACQSLRSGECTLALAGGVSVM 687
Db       1588 LTGNTASVMSGRVSYTLGLEGPALTVDTACSSSLVALHLAVQALRKGEVDMALAGGVAVM 1647

Qy        688 STLGMFIEFSRQRGLSVDGRCKAYSAAADGTGWGEGVGMLLVERLSDAVRLGHRVLAVVR 747
Db       1648 PTPGMFVEFSRQRGLAGDGRSKAFAASADGTSWSEGVGVLLVERLSDARRNGHQVLAVVR 1707

Qy        748 GSAVNQDGASNGLTAPNGPAQERVIRQALANAGLSVADVDVVEGHGTGTTLGDPIEAQAL 807
Db       1708 GSALNQDGASNGLTAPNGPSQQRVIRRALADARLTTSDVDVVEAHGTGTRLGDPIEAQAL 1767

Qy        808 LATYGQ-RAGDRPLWLGSLKSNIGHTMAAAGVGGVIKMVMALREGVLPRTLHVDKPSPQV 866
Db       1768 IATYGQGRDDEQPLRLGSLKSNIGHTQAAAGVSGVIKMVQAMRHGLLPKTLHVDEPSDQI 1827

Qy        867 DWSAGAVRLLTEAVPWPGDAAGRLRRAGVSSFGIGGTNAHVILEEAPAAGGCVAGGGVLE 926
Db       1828 DWSAGAVELLTEAVDWPEKQDGGLRRAAVSSFGISGTNAHVVLEEAPV---------VVE 1878

Qy        927 GAPGLAISVAESVAAPVAVSAPVAESVPVPVPVPVPVPVSARSEAGLRAQAEALRQYVAV 986
Db       1879 GA-----SVVEPSVGGSAVGGGV-----------TPWVVSAKSAAALDAQIERLAAF-AS 1921

Qy        987 RPDVSLADVGA--------GLACGRAVLEHRAVVLAADREELVQGLGALAAGEPDRRVTT 1038
Db       1922 RDRTDDADAGAVDAGAVAHVLADGRAQFEHRAVALGAGADDLVQAL-----ADPDGLI-R 1975

Qy       1039 GHAPGGDRGGVVFVFPGQGGQWAGMGVRLLASSPVFARRMQACEEALAPWVDWSVVDILR 1098
Db       1976 GTASG--VGRVAFVFPGQGTQWAGMGAELLDSSAVFAAAMAECEAALSPYVDWSLEAVVR 2033

Qy       1099 RDAGDAVWERADVVQPVLFSVMVSLAALWRSYGIEPDAVLGHSQGEIAAAHVCGALSLKD 1158
Db       2034 QAPGAPTLERVDVVQPVTFAVMVSLARVWQHHGVTPQAVVGHSQGEIAAAYVAGALPLDD 2093

Qy       1159 AAKTVALRSRALAA-VRGRGGMASVPLPAQEVEQLIGERWAGRLWVAAVNGPRSTAVSGD 1217
Db       2094 AARVVTLRSKSIAAHLAGKGGMLSLALNEDAVLERLSD--FDGLSVAAVNGPTATVVSGD 2151

Qy       1218 AEAVDEVLAYCAGTGVRARRIPVDYASHCPHVQPLREELLELLGDISPQPSGVPFFSTVE 1277
Db       2152 PVQIEELAQACKADGFRARIIPVDYASHSRQVEIIESELAQVLAGLSPQAPRVPFFSTLE 2211

Qy       1278 GTWLDTTTLDAAYWYRNLHQPVRFSDAVQALA-DDGHRVFVEVSPHPTLVPAIEDTTEDT 1336
Db       2212 GTWITEPVLDGTYWYRNLRHRVGFAPAIETLAVDEGFTHFVEVSAHPVLTMTLPET---- 2267

Qy       1337 AEDVTAIGSLRRGDNDTRRFLTALAHTHTTGIGTPTTWHHHYTHHHTHPHPHTHLDLPTY 1396
Db       2268 ---VTGLGTLRREQGGQERLVTSLAEAWVNGL--PVAWTSLLPATASRP------GLPTY 2316

Qy       1397 PFQHQHYWLESSQPGAGSGSGAGAGSGAGSGRAGTAGGTAEVESRFWDAVARQDLETVAT 1456
Db       2317 AFQAERYWLENTPAALATG---------------------------------------- 2335

Qy       1457 TLAVPPSAGLDTVVPALSAWHRHQHDQARINTWTYQETWKPLTLPTTHQPHQT-----WL 1511
Db       2336 ------------------------------DDWRYRIDWK--RLPAAEGSERTGLSGRWL 2363

Qy       1512 IAIPETQTHHPHITNILTNLHHHGITPIPLTLNHTHTNPQHLHHTLHHTRQQAQNHTTG- 1570
Db       2364 AVTPE--DHSAQAAAVLTALVDAGAKVEVLTAG-ADDDREALAARLTAL-------TTGD 2413

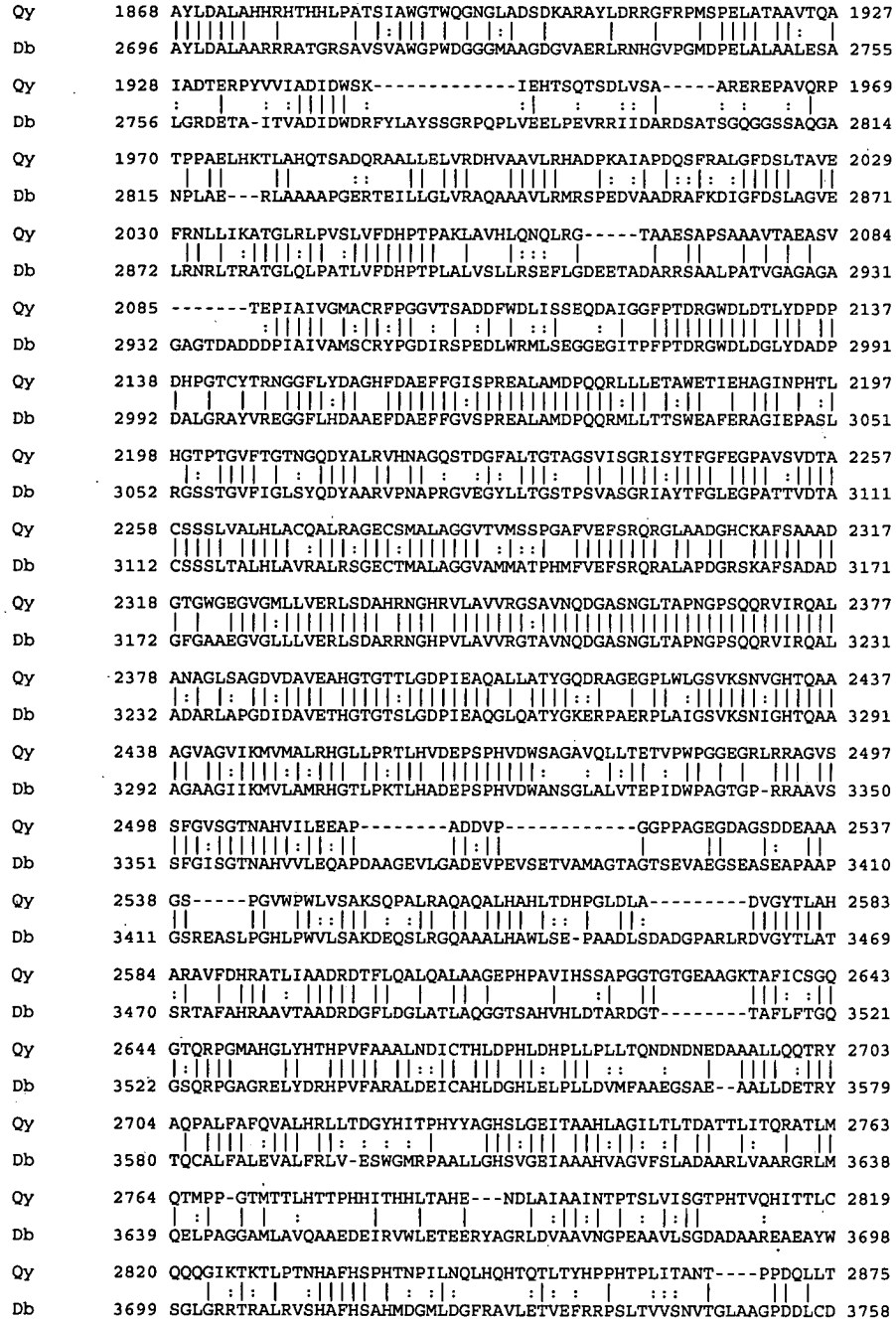
Qy       1571 AITGLLSLLALDETPHPHHPHTPTGTLLNLTLTQTHTQTHPPTPLWYATTNATTTHPNDP 1630
Db       2414 GFTGVVSLL-------------DGLVPQVAWVQALGDAGIKAPLWSVTQGAVSVGRLDT 2459

Qy       1631 LTHPTQAQTWGLARTTLLEHPTHTAGIIDLPTTPTPHTLQHLTQTLTQPHHQTQLAIRTT 1690
Db       2460 PADPDRAMLWGLGRVVALEHPERWAGLVDLPAQPDAAALAHLVTALSGATGEDQIAIRTT 2519

Qy       1691 GTHTRRLTPTTLTPTH-QPPTP--TPHGTTLITGGTGALATHLTHHLTTHQPTQHLLLTS 1747
Db       2520 GLHARRL---ARAPLHGRRPTRDWQPHGTVLITGGTGALGSHAARWM-AHHGAEHLLLVS 2575

Qy       1748 RTGPHTPHAQHLTTQLQQKGIHLTITTCDTSNPDQLQQLLNTIPPQHPLTTVIHTAGILD 1807
Db       2576 RSGEQAPGATQLTAELTASGARVTIAACDVADPHAMRTLLDAIPAETPLTAVVHTAGALD 2635

Qy       1808 DATLTNLTPTQLNNVLRAKAHSAHLLHQLTQHTPLTAFVLYSSAAATFGAPGQANYAAAN 1867
Db       2636 DGIVDTLTAEQVRRAHRAKAVGASVLDELTRDLDLDAFVLFSSVSSTLGIPGQGNYAPHN 2695
```


Qy 1868 AYLDALAHHRHTHHLPATSIAWGTWQGNGLADSDKARAYLDRRGFRPMSPELATAAVTQA 1927
Db 2696 AYLDALAARRRATGRSAVSVAWGPWDGGGMAAGDGVAERLRNHGVPGMDPELALAALESA 2755
Qy 1928 IADTERPYVVIADIDWSK------------IEHTSQTSDLVSA-----ARELEPAVQRP 1969
Db 2756 LGRDETA-ITVADIDWDRFYLAYSSGRPQPLVEELPEVRRIIDARDSATSGQGGSSAQGA 2814
Qy 1970 TPPAELHKTLAHQTSADQRAALLELVRDHVAAVLRHADPKAIAPDQSFRALGFDSLTAVE 2029
Db 2815 NPLAE---RLAAAAPGERTEILLGLVRAQAAAVLRMRSPEDVAADRAFKDIGFDSLAGVE 2871
Qy 2030 FRNLLIKATGLRLPVSLVFDHPTPAKLAVHLQNQLRG-----TAAESAPSAAAVTAEASV 2084
Db 2872 LRNRLTRATGLQLPATLVFDHPTPLALVSLLRSEFLGDEETADARRSAALPATVGAGAGA 2931
Qy 2085 -------TEPIAIVGMACRFPGGVTSADDFWDLISSEQDAIGGFPTDRGWDLDTLYDPDP 2137
Db 2932 GAGTDADDDPIAIVAMSCRYPGDIRSPEDLWRMLSEGGEGITPFPTDRGWDLDGLYDADP 2991
Qy 2138 DHPGTCYTRNGGFLYDAGHFDAEFFGISPREALAMDPQQRLLLETAWETIEHAGINPHTL 2197
Db 2992 DALGRAYVREGGFLHDAAEFDAEFFGVSPREALAMDPQQRMLLTTSWEAFERAGIEPASL 3051
Qy 2198 HGTPTGVFTGTNGQDYALRVHNAGQSTDGFALTGTAGSVISGRISYTFGFEGPAVSVDTA 2257
Db 3052 RGSSTGVFIGLSYQDYAARVPNAPRGVEGYLLTGSTPSVASGRIAYTFGLEGPATTVDTA 3111
Qy 2258 CSSSLVALHLACQALRAGECSMALAGGVTVMSSPGAFVEFSRQRGLAADGHCKAFSAAAD 2317
Db 3112 CSSSLTALHLAVRALRSGECTMALAGGVAMMATPHMFVEFSRQRALAPDGRSKAFSADAD 3171
Qy 2318 GTGWGEGVGMLLVERLSDAHRNGHRVLAVVRGSAVNQDGASNGLTAPNGPSQQRVIRQAL 2377
Db 3172 GFGAAEGVGLLLVERLSDARRNGHPVLAVVRGTAVNQDGASNGLTAPNGPSQQRVIRQAL 3231
Qy 2378 ANAGLSAGDVDAVEAHGTGTTLGDPIEAQALLATYGQDRAGEGPLWLGSVKSNVGHTQAA 2437
Db 3232 ADARLAPGDIDAVETHGTGTSLGDPIEAQGLQATYGKERPAERPLAIGSVKSNIGHTQAA 3291
Qy 2438 AGVAGVIKMVMALRHGLLPRTLHVDEPSPHVDWSAGAVQLLTETVPWPGGEGRLRRAGVS 2497
Db 3292 AGAAGIIKMVLAMRHGTLPKTLHADEPSPHVDWANSGLALVTEPIDWPAGTGP-RRAAVS 3350
Qy 2498 SFGVSGTNAHVILEEAP--------ADDVP-----------GGPPAGEGDAGSDDEAAA 2537
Db 3351 SFGISGTNAHVVLEQAPDAAGEVLGADEVPEVSETVAMAGTAGTSEVAEGSEASEAPAAP 3410
Qy 2538 GS-----PGVWPWLVSAKSQPALRAQAQALHAHLTDHPGLDLA---------DVGYTLAH 2583
Db 3411 GSREASLPGHLPWVLSAKDEQSLRGQAAALHAWLSE-PAADLSDADGPARLRDVGYTLAT 3469
Qy 2584 ARAVFDHRATLIAADRDTFLQALQALAAGEPHPAVIHSSAPGGTGTGEAAGKTAFICSGQ 2643
Db 3470 SRTAFAHRAAVTAADRDGFLDGLATLAQGGTSAHVHLDTARDGT--------TAFLFTGQ 3521
Qy 2644 GTQRPGMAHGLYHTHPVFAAALNDICTHLDPHLDHPLLPLLTQNDNDNEDAAALLQQTRY 2703
Db 3522 GSQRPGAGRELYDRHPVFARALDEICAHLDGHLELPLLDVMFAAEGSAE--AALLDETRY 3579
Qy 2704 AQPALFAFQVALHRLLTDGYHITPHYYAGHSLGEITAAHLAGILTLTDATTLITQRATLM 2763
Db 3580 TQCALFALEVALFRLV-ESWGMRPAALLGHSVGEIAAAHVAGVFSLADAARLVAARGRLM 3638
Qy 2764 QTMPP-GTMTTLHTTPHHITHHLTAHE---NDLAIAAINTPTSLVISGTPHTVQHITTLC 2819
Db 3639 QELPAGGAMLAVQAAEDEIRVWLETEERYAGRLDVAAVNGPEAAVLSGDADAAREAEAYW 3698
Qy 2820 QQQGIKTKTLPTNHAFHSPHTNPILNQLHQHTQTLTYHPPHTPLITANT----PPDQLLT 2875
Db 3699 SGLGRRTRALRVSHAFHSAHMDGMLDGFRAVLETVEFRRPSLTVVSNVTGLAAGPDDLCD 3758
Qy 2876 PHYWTQQARNTVDYATTTQTLHQHGVTTYIELGPDNTLTTLTHHNLPNPPTTTLT---LT 2932
Db 3759 PEYWVRHVRGTVRFLDGVRVLRDLGVRTCLELGPDGVLTAMAADGLADTPADSAAGSPVG 3818
Qy 2933 HPHHHPQTH---------LLTNL-----AKTTTTWHPHHYTH-HDNQPHTHT-------- 2969
Db 3819 SPAGSPADSAAGALRPRPLLVALLRRKRSETETVADALGRAHAHGTGPDWHAWFAGSGAH 3878
Qy 2970 HLDLPTYPFQHHHYWLESTQPGAG---NVSAAGLDPTEHPLLGATLELATDGGALLAGRL 3026
Db 3879 RVDLPTYSFRRDRYWLDA--PAADTAVDTAGLGLGTADHPLLGAVVSLPDRDGLLLTGRL 3936
Qy 3027 SLRSHPWLADHAVGGTVLLSGATFLELALHAGTYVGCDRVDELTLHAPLVVPVDGGVSVQ 3086
Db 3937 SLRTHPWLADHAVLGSVLLPGAAMVELAAHAAESAGLRDVRELTLLEPLVLPEHGGVELR 3996
Qy 3087 VGVAAADGE--------GRRLVSVYARGGSACGGGGASGGVWTCHASGVLVEAAAGGVVV 3138
Db 3997 VTVGAPAGEPGGESAGDGARPVSLHSRLADA-----PAGTAWSCHATGLLATDRPELPVA 4051
Qy 3139 DGLAGVWPPRGAVAVDVDGVRDRLAGAGCVLGPVFSGLRAVWRDGGDLLAEVCLP----- 3193
Db 4052 PDRAAMWPPQGAEEVPLDGLYERLDGNGLAFGPLFQGLNAVWRYEGEVFADIALPATTNA 4111
Qy 3194 -----EEAWGDAAG--FGLHPALLDGVVQPLSVLLPGGTGFGEGAGFGEGVRVPAVWGGV 3246
Db 4112 TAPATANGGGSAAAAPYGIHPALLDASLHAIAV---GGL-----VDEPELVRVPFHWSGV 4163
Qy 3247 SLHRAGVTGVRVRVSAVGRGGGREAVSVVVGDEAGVPVASVDRLELRPVDMGQLRAVSVS 3306
Db 4164 TVHAAGAAAARVRLAS----AGTDAVSLSLTDGEGRPLVSVERLTLRPVTADQAAASRVG 4219
Qy 3307 AGRRGSLYAVQW---------------AEVGPVPVCGQ---------------------A 3330
Db 4220 ----GLMHRVAWRPYALASSGEQDPHATSYGPTAVLGKDELKVAAALESAGVEVGLYPDL 4275
Qy 3331 WAWHEDVGESGGGPVPGVVVLRCPDAGAGGGGGGGGGGVGEVVGGVLGVVQGWLGLERF 3390
Db 4276 AALSQDV--AAGAPAPRTVLAPLP-----AGPADGGAEGVRGTVARTLELLQAWLADEHL 4328
Qy 3391 AGSRLVVVTRGAV----VAGPEDGPVDVVGASVWGLVRSAQAEHPDRFVLLDLDTDTGT- 3445
Db 4329 AGTRLLLVTRGAVRDPEGSGADDGGEDLSHAAAWGLVRTAQTENPGRFGLLDLADDASSY 4388
Qy 3446 ----DLDTGAGAGWGVDGGRVAAVVACGEPQLAVRGERLLAARLKRLESSGDVPAQRSGD 3501
Db 4389 RTLPSVLSDAGLR--------------DEPQLALHDGTIRLARL---------------- 4418
Qy 3502 TRARRSDVPAQRSGGVPARRSVDVSGREVLPWLSGGSVLVTGGTGVLGAAVARHLAGVCG 3561
Db 4419 -----ASVRPETGTAAPA----------LAP---EGTVLLTGGTGGLGGLVARHVVGEWG 4460
Qy 3562 VRDLLLVSRRGPDAPGAEGLRAELAALGAEVRIVACDVGERREVVRLLEGVPAGCPLTGV 3621
Db 4461 VRRLLLVSRRGTDAPGADELVHELEALGADVSVAACDVADREALTAVLDAIPAEHPLTAV 4520
Qy 3622 VHAAGVLDDATIASLTPERLGTVFAAKVDAALLLDELTR--GMELSAFVLFSSAAGILGS 3679
Db 4521 VHTAGVLSDGTLPSMTTEDVEHVLRPKVDAAFLLDELTSTPAYDLAAFVMFSSAAAVFGG 4580
Qy 3680 AGQGNYAAANAALDALAYRRRAAGLPGVSLAWGLWEEASGMTGHLAGTDHRRIIRSGLHP 3739
Db 4581 AGQGAYAAANATLDALAWRRRAAGLPALSLGWGLWAETSGMTGELGQADLRRMSRAGIGG 4640
Qy 3740 MSTPDALALFDAALALDR-PVLLP-----ADLRPA-----PPLPPLLQDLLPATRRR--- 3785
Db 4641 ISDAEGIALLDAALRDDRHPVLLPLRLDAAGLRDAAGNDPAGIPALFRDVVGARTVRARP 4700
Qy 3786 ------TTRTT--TTGGADNGAQLHA-----RLAGQTHEQQHTTLLALVRSHIATVLGHT 3832
Db 4701 SAASASTTAGTAGTPGTADGAAETAAVTLADRAATVDGPARQRLLLEFVVGEVAEVLGHA 4760
Qy 3833 TPDTIPPDRAFRDLGFDSLTAVELRNRLSRTTGLRLPTTLAFDHPNPTTLTHHLHTQL-- 3890

```
Db       4761 RGHRIDAERGFLDLGFDSLTAVELRNRLNSAGGLALPATLVFDHPSPAALASHLDAELPR 4820

Qy       3891 -QPQPDNA------------------VAPVLAELDKLESALSALDKTDS-ASERVTLRLK 3930

Db       4821 GASDQDGAGNRNGNENGTTASRSTAETDALLAQLTRLEGALVLTGLSDAPGSEEVLEHLR 4880

Qy       3931 SLMLR------------------------WNAPQHPTAESADDD--EKFTSATEAEIFK 3963

Db       4881 SLRSMVTGETGTGTASGAPDGAGSGAEDRPWAAGDGAGGGSEDGAGVPDFMNASAEELFG 4940

Qy       3964 FIDND 3968

Db       4941 LLDQD 4945


RESULT 9
US-09-988-384B-6
; Sequence 6, Application US/09988384B
; Publication No. US20030073824A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.536US1
;  CURRENT APPLICATION NUMBER: US/09/988,384B
;  CURRENT FILING DATE:  2001-11-19
;  PRIOR APPLICATION NUMBER: PCT/US99/14398
;  PRIOR FILING DATE: 1999-06-25
;  PRIOR APPLICATION NUMBER: US 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 53
; SEQ ID NO 6
;   LENGTH: 12199
;   TYPE: PRT
;   ORGANISM: Streptomyces venezuelae
US-09-988-384B-6

  Query Match             39.1%;  Score 8148;  DB 9;  Length 12199;
  Best Local Similarity   44.8%;  Pred. No. 0;
  Matches 1909;  Conservative  463;  Mismatches 1307;  Indels  586;  Gaps   94;

Qy         15 GEVLGVADEADGGVVFVFPGQGPQWPGMGRELLDASDVFRESVRACEAAFAPYVDWSVEQ 74

Db       1278 GLVRGVASGV-GRVAFVFPGQGTQWAGMGAELLDSSAVFAAAMAECEAALSPYVDWSLEA 1336

Qy         75 VLRDSPDAPGLDRVDVVQPTLFAVMISLAALWRSQGVEPCAVLGHSLGEIAAAHVSGGLS 134

Db       1337 VVRQAPGAPTLERVDVVQPVTFAVMVSLARVWQHHGVTPQAVVGHSQGEIAAAYVAGALS 1396

Qy        135 LADAARVVTLWSQA-QTTLAGTGALVSVAATPDELLPRIAPWTEDNPARLAVAAVNGPRS 193

Db       1397 LDDAARVVTLRSKSIAAHLAGKGGMLSLALSEDAVLERLAGFD-----GLSVAAVNGPTA 1451

Qy        194 TVVSGAREAVADLVADLTAAQVRTRMIPVDVPAHSPLMYAIEERVVSGLLPITPRPSRIP 253

Db       1452 TVVSGDPVQIEELARACEADGVRARVIPVDYASHSRQVEIIESELAEVLAGLSPQAPRVP 1511

Qy        254 FHSSVTGGRLDTRELDAAYWYRNMSSTVRFEPAAR-LLLQQGPKTFVEMSPHPVLTMGLQ 312

Db       1512 FFSTLEGAWITEPVLDGGYWYRNLRHRVGFAPAVETLATDEGFTHPVEVSAHPVLTMALP 1571

Qy        313 ELAPDLGDTTGTADTVIMGTLRRGQGTLDHPLTSLAQ---------------LRGHGET 356

Db       1572 ------GTVTGLA------TLRRDNGGQDRLVASLAEAWANGLAVDWSPLLPSATGHHSD 1619

Qy        357 SATTVLS------ARLTALSPT----------------------QQQSLLLDLVRAHTM 387

Db       1620 LPTYAFQTERHWLGEIEALAPAGEPAVQPAVLRTEAAEPAELDRDEQLRVILDKVRAQTA 1679

Qy        388 AVLNDDGNERTASDAGPSASFAHLGFDSVMGVELRNRLSKATGLRLPVTLIFDHTTPAAV 447

Db       1680 QVL----GYATGGQIEVDRTFREAGCTSLTGVDLRNRINAAFGVRMAPSMIFDFPTPEAL 1735

Qy        448 AARLRTAALGHLDEDTAPVPDSPSGHGGTAAADDPIAIIGMACRFPGGVRSPKDLWELAA 507

Db       1736 AEQLLLVVHGEAAANPAGAEPAPVAAAG--AVDEPVAIVGMACRLPGGVASPEDLWRLVA 1793

Qy        508 SGGDAIGPFPTDRGWPTEQRHAQDPTQPGTFYPQGGGFLHDAAHFDAGFFGISPREALAM 567

Db       1794 GGGDAISEFPQDRGWDVEGLYHPDPEHPGTSYVRQGGFIENVAGFDAAFFGISPREALAM 1853

Qy        568 DPQQRLLLETSWEAFERAGIDPLSVRGSRTGVFAGALSFDYGPRMDTASSEGAADVEGHI 627

Db       1854 DPQQRLLLETSWEAVEDAGIDPTSLRGRQVGVFTGAMTHEYGPSL----RDGGEGLDGYL 1909

Qy        628 LTGTTGSVLSGRIAYSFGLEGPAITVDTGCSASLVTLHLACQSLRSGECTLALAGGVSVM 687

Db       1910 LTGNTASVMSGRVSYTLGLEGPALTVDTACSSSLVALHLAVQALRKGEVDMALAGGVAVM 1969

Qy        688 STLGMFIEFSRQRGLSVDGRCKAYSAAADGTGWGEGVGMLLVERLSDAVRLGHRVLAVVR 747

Db       1970 PTPGMFVEFSRQRGLAGDGRSKAFAASADGTSWSEGVGVLLVERLSDARRNGHQVLAVVR 2029

Qy        748 GSAVNQDGASNGLTAPNGPAQERVIRQALANAGLSVADVDVVEGHGTGTTLGDPIEAQAL 807

Db       2030 GSALNQDGASNGLTAPNGPSQQRVIRRALADARLTTSDVDVVEAHGTGTRLGDPIEAQAL 2089

Qy        808 LATYGQ-RAGDRPLWLGSLKSNIGHTMAAAGVGGVIKMVMALREGVLPRTLHVDKPSPQV 866

Db       2090 IATYGQGRDDEQPLRLGSLKSNIGHTQAAAGVSGVIKMVQAMRHGLLPKTLHVDEPSDQI 2149

Qy        867 DWSAGAVRLLTEAVPWPGDAAGRLRRAGVSSFGIGGTNAHVILEEAPAAGGCVAGGGVLE 926

Db       2150 DWSAGAVELLTEAVDWPEKQDGGLRRAAVSSFGISGTNAHVVLEEAPV---------VVE 2200

Qy        927 GAPGLAISVAESVAAPVAVSAPVAESVPVPVPVPVPVPVSARSEAGLRAQAEALRQYVAV 986

Db       2201 GA-----SVVEPSVGGSAVGGGV----------TPWVVSAKSAAALDAQIERLAAF-AS 2243

Qy        987 RPDVSLADVGA--------GLACGRAVLEHRAVVLAADREELVQGLGALAAGEPDRRVTT 1038

Db       2244 RDRTDDADAGAVDAGAVAHVLADGRAQFEHRAVALGAGADDLVQAL-----ADPDGLI-R 2297

Qy       1039 GHAPGGDRGGVVFVFPGQGGQWAGMGVRLLASSPVFARRMQACEEALAPWVDWSVVDILR 1098

Db       2298 GTASG--VGRVAFVFPGQGTQWAGMGAELLDSSAVFAAAMAECEAALSPYVDWSLEAVVR 2355

Qy       1099 RDAGDAVWERADVVQPVLFSVMVSLAALWRSYGIEPDAVLGHSQGEIAAAHVCGALSLKD 1158

Db       2356 QAPGAPTLERVDVVQPVTFAVMVSLARVWQHHGVTPQAVVGHSQGEIAAAYVAGALPLDD 2415

Qy       1159 AAKTVALRSRALAA-VRGRGGMASVPLPAQEVEQLIGERWAGRLWVAAVNGPRSTAVSGD 1217

Db       2416 AARVVTLRSKSIAAHLAGKGGMLSLALNEDAVLERLSD--FDGLSVAAVNGPTATVVSGD 2473

Qy       1218 AEAVDEVLAYCAGTGVRARRIPVDYASHCPHVQPLREELLELLGDISPQPSGVPFFSTVE 1277

Db       2474 PVQIEELAQACKADGFRARIIPVDYASHSRQVEIIESELAQVLAGLSPQAPRVPFFSTLE 2533

Qy       1278 GTWLDTTTLDAAYWYRNLHQPVRFSDAVQALA-DDGHRVFVEVSPHPTLVPAIEDTTEDT 1336

Db       2534 GTWITEPVLDGTYWYRNLRHRVGFAPAIETLAVDEGFTHPVEVSAHPVLTMTLPET---- 2589

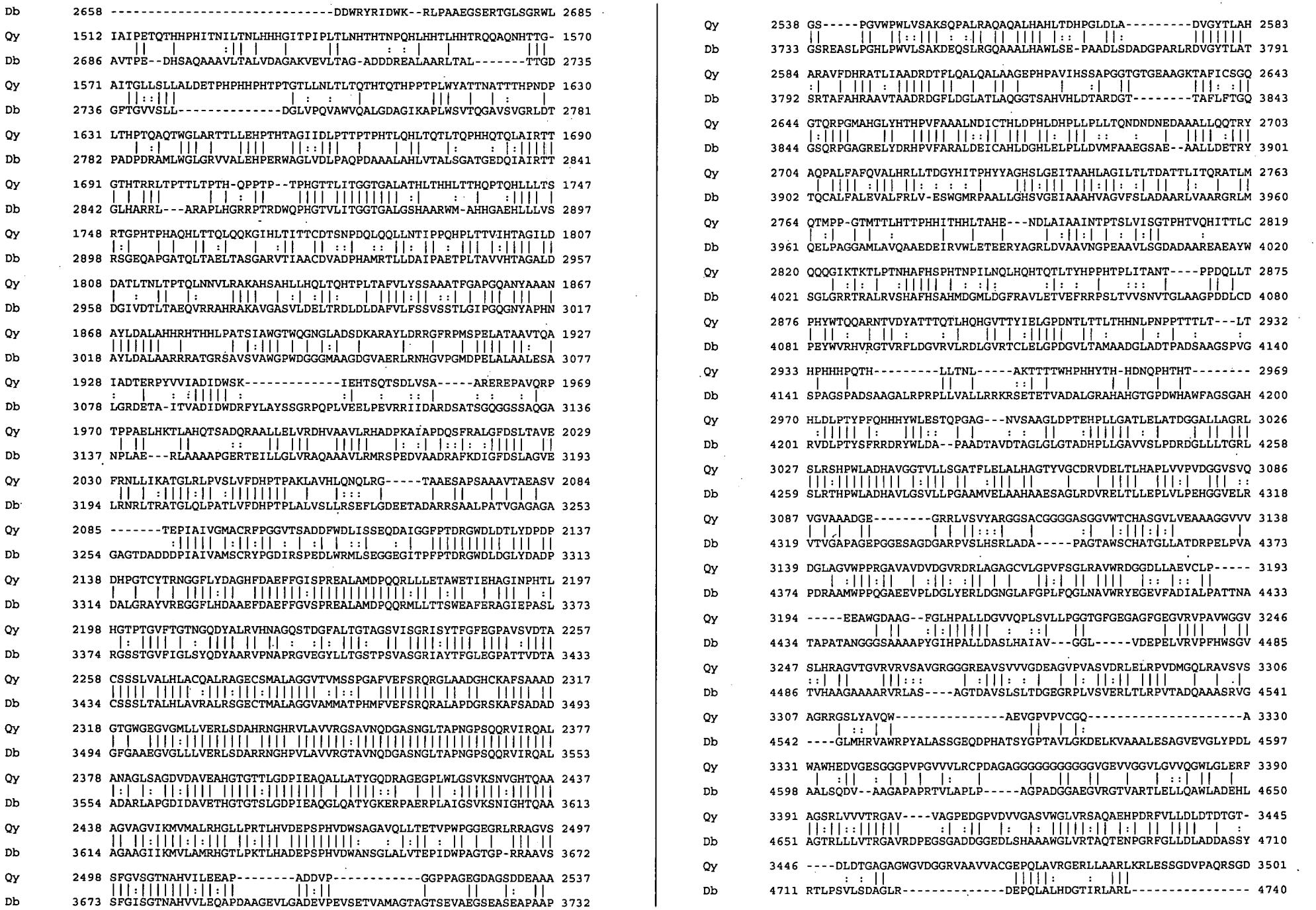
Qy       1337 AEDVTAIGSLRRGDNDTRRFLTALAHTHTTGIGTPTTWHHHYTHHHTHPHPHTHLDLPTY 1396

Db       2590 ---VTGLGTLRREQGGQERLVTSLAEAWVNGL--PVAWTSLLPATASRP------GLPTY 2638

Qy       1397 PFQHQHYWLESSQPGAGSGSGAGAGSGAGSGRAGTAGGTAEVESRFWDAVARQDLETVAT 1456

Db       2639 AFQAERYWLENTPAALATG---------------------------------------- 2657

Qy       1457 TLAVPPSAGLDTVVPALSAWHRHQHDQARINTWTYQETWKPLTLPTTHQPHQT-----WL 1511
```



```
Qy       3502 TRARRSDVPAQRSGGVPARRSVDVSGREVLPWLSGGSVLVTGGTGVLGAAVARHLAGVCG 3561
Db       4741 -----ASVRPETGTAAPA----------LAP---EGTVLLTGGTGGLGGLVARHVVGEWG 4782

Qy       3562 VRDLLLVSRRGPDAPGAEGLRAELAALGAEVRIVACDVGERREVVRLLEGVPAGCPLTGV 3621
Db       4783 VRRLLLVSRRGTDAPGADELVHELEALGADVSVAACDVADREALTAVLDAIPAEHPLTAV 4842

Qy       3622 VHAAGVLDDATIASLTPERLGTVFAAKVDAALLLDELTR--GMELSAFVLFSSAAGILGS 3679
Db       4843 VHTAGVLSDGTLPSMTTEDVEHVLRPKVDAAFLLDELTSTPAYDLAAFVMFSSAAAVFGG 4902

Qy       3680 AGQGNYAAANAALDALAYRRRAAGLPGVSLAWGLWEEASGMTGHLAGTDHRRIIRSGLHP 3739
Db       4903 AGQGAYAAANATLDALAWRRRAAGLPALSLGWGLWAETSGMTGELGQADLRRMSRAGIGG 4962

Qy       3740 MSTPDALALFDAALALDR-PVLLP-----ADLRPA-----PPLPPLLQDLLPATRRR--- 3785
Db       4963 ISDAEGIALLDAALRDDRHPVLLPLRLDAAGLRDAAGNDPAGIPALFRDVVGARTVRARP 5022

Qy       3786 ------TTRTT--TTGGADNGAQLHA-----RLAGQTHEQQHTTLLALVRSHIATVLGHT 3832
Db       5023 SAASASTTAGTAGTPGTADGAAETAAVTLADRAATVDGPARQRLLLEFVVGEVAEVLGHA 5082

Qy       3833 TPDTIPPDRAFRDLGFDSLTAVELRNRLSRTTGLRLPTTLAFDHPNPTTLTHHLHTQL-- 3890
Db       5083 RGHRIDAERGFLDLGFDSLTAVELRNRLNSAGGLALPATLVFDHPSPAALASHLDAELPR 5142

Qy       3891 -QPQPDNA-----------------VAPVLAELDKLESALSALDKTDS-ASERVTLRLK 3930
Db       5143 GASDQDGAGNRNGNENGTTASRSTAETDALLAQLTRLEGALVLTGLSDAPGSEEVLEHLR 5202

Qy       3931 SLMLR-----------------------WNAPQHPTAESADDD--EKPTSATEAEIFK 3963
Db       5203 SLRSMVTGETGTGTASGAPDGAGSGAEDRPWAAGDGAGGGSEDGAGVPDFMNASAEELFG 5262

Qy       3964 FIDND 3968
Db       5263 LLDQD 5267
```

```
RESULT 10
US-09-793-708-2
; Sequence 2, Application US/09793708
; Publication No. US20030104597A1
; GENERAL INFORMATION:
;  APPLICANT: ASHLEY, Gary
;  APPLICANT:  BETLACH, Melanie C.
;  APPLICANT:  BETLACH, Mary C.
;  APPLICANT:  McDANIEL, Robert
;  APPLICANT:  TANG, Li
;  TITLE OF INVENTION: RECOMBINANT NARBONOLIDE POLYKETIDE SYNTHASE
;  FILE REFERENCE: 300622002121
;  CURRENT APPLICATION NUMBER: US/09/793,708
;  CURRENT FILING DATE:  2001-02-22
;  PRIOR APPLICATION NUMBER: US 09/657,440
;  PRIOR FILING DATE: 2000-09-07
;  PRIOR APPLICATION NUMBER: US 09/320,878
;  PRIOR FILING DATE: 1999-05-27
;  PRIOR APPLICATION NUMBER: US 09/141,908
;  PRIOR FILING DATE: 1998-08-28
;  PRIOR APPLICATION NUMBER: US 09/073,538
;  PRIOR FILING DATE: 1998-05-06
;  PRIOR APPLICATION NUMBER: US 08/846,247
;  PRIOR FILING DATE: 1997-04-30
;  PRIOR APPLICATION NUMBER: US 60/134,990
;  PRIOR FILING DATE: 1999-05-20
;  NUMBER OF SEQ ID NOS: 38
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
;   LENGTH: 3739
;   TYPE: PRT
;   ORGANISM: Streptomyces venezuelae
```

US-09-793-708-2

Query Match 34.0%; Score 7079; DB 9; Length 3739;
Best Local Similarity 42.5%; Pred. No. 0;
Matches 1681; Conservative 391; Mismatches 1171; Indels 712; Gaps 81;

```
Qy        477 AAADDPIAIIGMACRFPGGVRSPKDLWELAASGGDAIGPFPTDRGWPTEQRHAQDPTQPG 536
Db         32 AKAGEPVAIVGMACRLPGGVASPEDLWRLVAGGEDAISEFPQDRGWDVEGLYDPNPEATG 91

Qy        537 TFYPQGGGFLHDAAHFDAGFFGISPREALAMDPQQRLLLETSWEAFERAGIDPLSVRGSR 596
Db         92 KSYAREAGFLYEAGEFDADFFGISPREALAMDPQQRLLLEASWEAFEHAGIPAATARGTS 151

Qy        597 TGVFAGALSFDYGPRMDTASSEGAADVEGHILTGTTGSVLSGRIAYSFGLEGPAITVDTG 656
Db        152 VGVFTGVMYHDYATRL-TDVPEG---IEGYLGTGNSGSVASGRVAYTLGLEGPAVTVDTA 207

Qy        657 CSASLVTLHLACQSLRSGECTLALAGGVSVMSTLGMFIEFSRQRGLSVDGRCKAYSAAAD 716
Db        208 CSSSLVALHLAVQALRKGEVDMALAGGVTVMSTPSTFVEFSRQRGLAPDGRSKSFSSTAD 267

Qy        717 GTGWGEGVGMLLVERLSDAVRLGHRVLAVVRGSAVNQDGASNGLTAPNGPAQERVIRQAL 776
Db        268 GTSWSEGVGVLLVERLSDARRKGHRILAVVRGTAVNQDGASSGLTAPNGPSQQRVIRRAL 327

Qy        777 ANAGLSVADVDVVEGHGTGTTLGDPIEAQALLATYGQ-RAGDRPLWLGSLKSNIGHTMAA 835
Db        328 ADARLTTSDVDVVEAHGTGTRLGDPIEAQAVIATYGQGRDGEQPLRLGSLKSNIGHTQAA 387

Qy        836 AGVGGVIKMVMALREGVLPRTLHVDKPSPQVDWSAGAVRLLTEAVPWPGDAAGRLRRAGV 895
Db        388 AGVSGVIKMVQAMRHGVLPKTLHVEKPTDQVDWSAGAVELLTEAMDWPDKGDGGLRRAAV 447

Qy        896 SSFGIGGTNAHVILEEAPAAGGCVAGGGVLEGAPGLAISVAESVAAPVAVSAPVAESVPV 955
Db        448 SSFGVSGTNAHVVLEEAPA----------------------------AEETPASEATPA 478

Qy        956 PVPVP----VPVPVSARSEAGLRAQAEALRQYVAVRPDVSLADVGA---GLACGRAVLEH 1008
Db        479 VEPSVGAGLVPWLVSAKTPAALDAQIGRLAAF-ASQGRTDAADPGAVARVLAGGRAEFEH 537

Qy       1009 RAVVLAADREELVQGLGA---LAAGEPDRRVTTGHAPGGDRGGVVFVFPGQGGQWAGMGV 1065
Db        538 RAVVLGTGQDDFAQALTAPEGLIRGTP----------SDVGRVAFVFPGQGTQWAGMGA 586

Qy       1066 RLLASSPVFARRMQACEEALAPWVDWSVVDILRRDAGDAVWERADVVQPVLFSVMVSLAA 1125
Db        587 ELLDVSKEFAAAMAECESALSRYVDWSLEAVVRQAPGAPTLERVDVVQPVTFAVMVSLAK 646

Qy       1126 LWRSYGIEPDAVLGHSQGEIAAAHVCGALSLKDAAKTVALRSRALAA-VRGRGGMASVPL 1184
Db        647 VWQHHGVTPQAVVGHSQGEIAAAYVAGALTLDDAARVVTLRSKSIAAHLAGKGGMISLAL 706

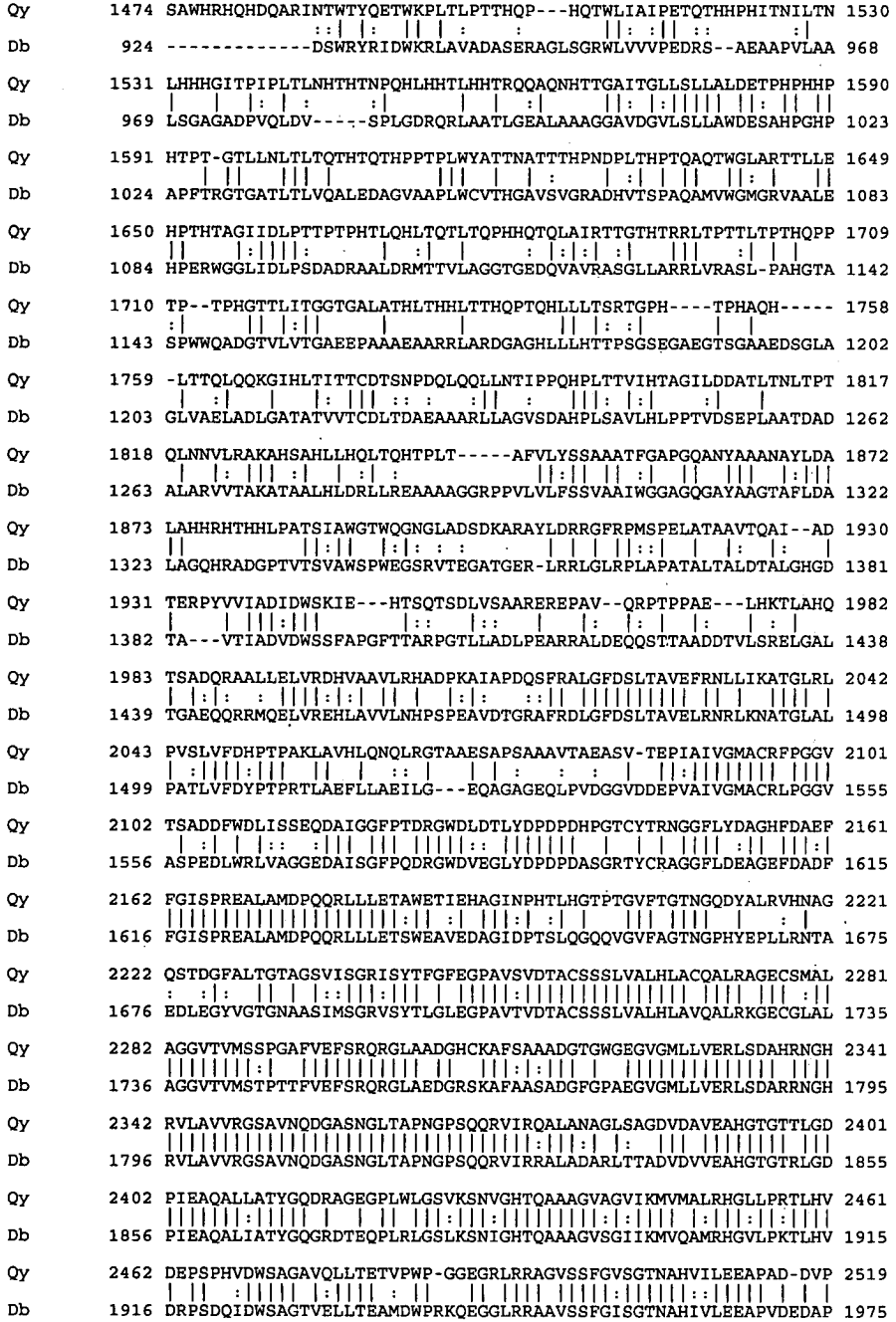
Qy       1185 ----PAQEVEQLIGERWAGRLWVAAVNGPRSTAVSGDAEAVDEVLAYCAGTGVRARRIPV 1240
Db        707 SEEATRQRIENLHG------LSIAAVNGPTATVVSGDPTQIQELAQACEADGVRARIIPV 760

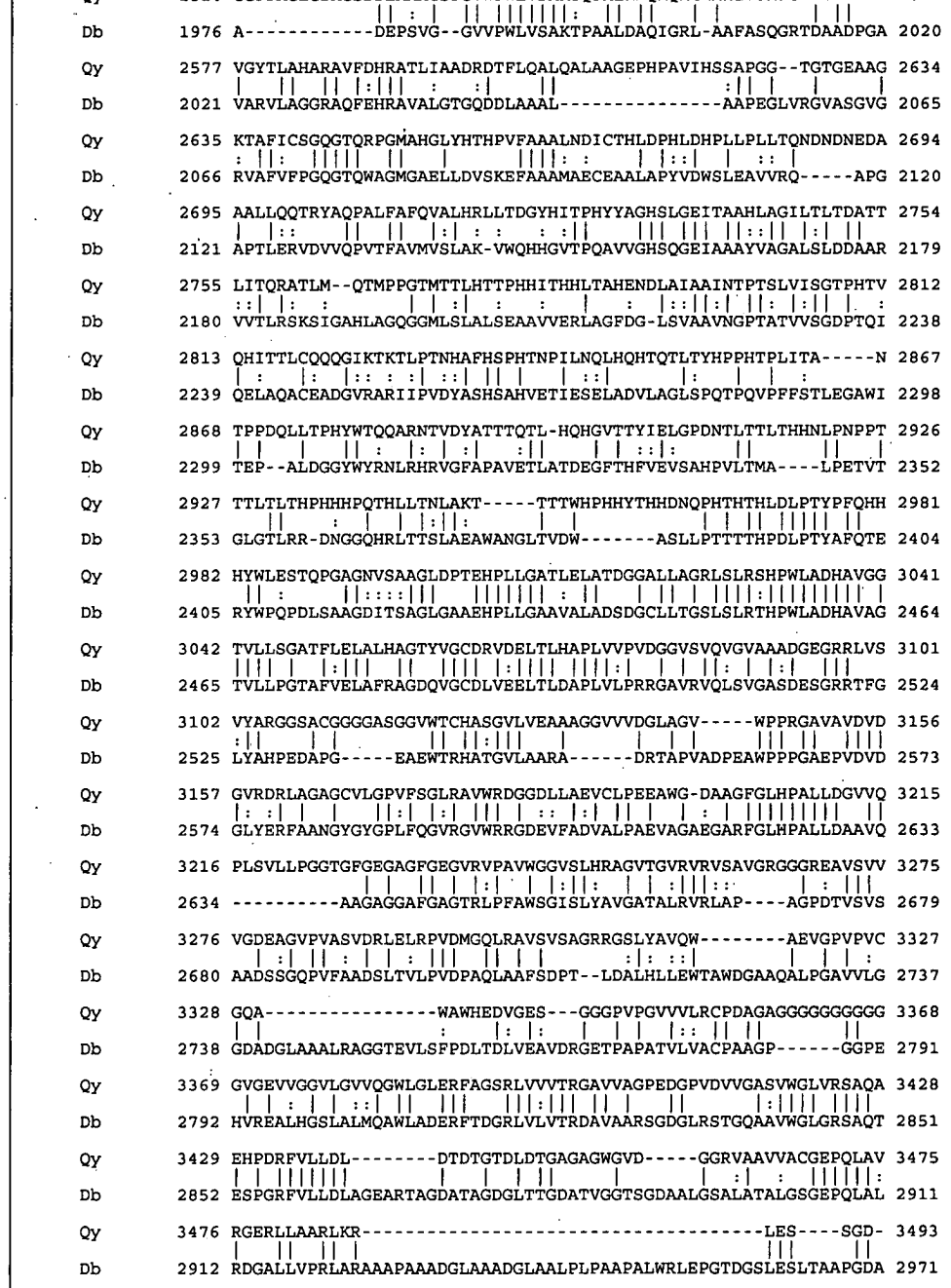
Qy       1241 DYASHCPHVQPLREELLELLGDISPQPSGVPFFSTVEGTWLDTTTLDAAYWYRNLHQPVR 1300
Db        761 DYASHSAHVETIESELAEVLAGLSPRTPEVPFFSTLEGAWITEPVLDGTYWYRNLRHRVG 820

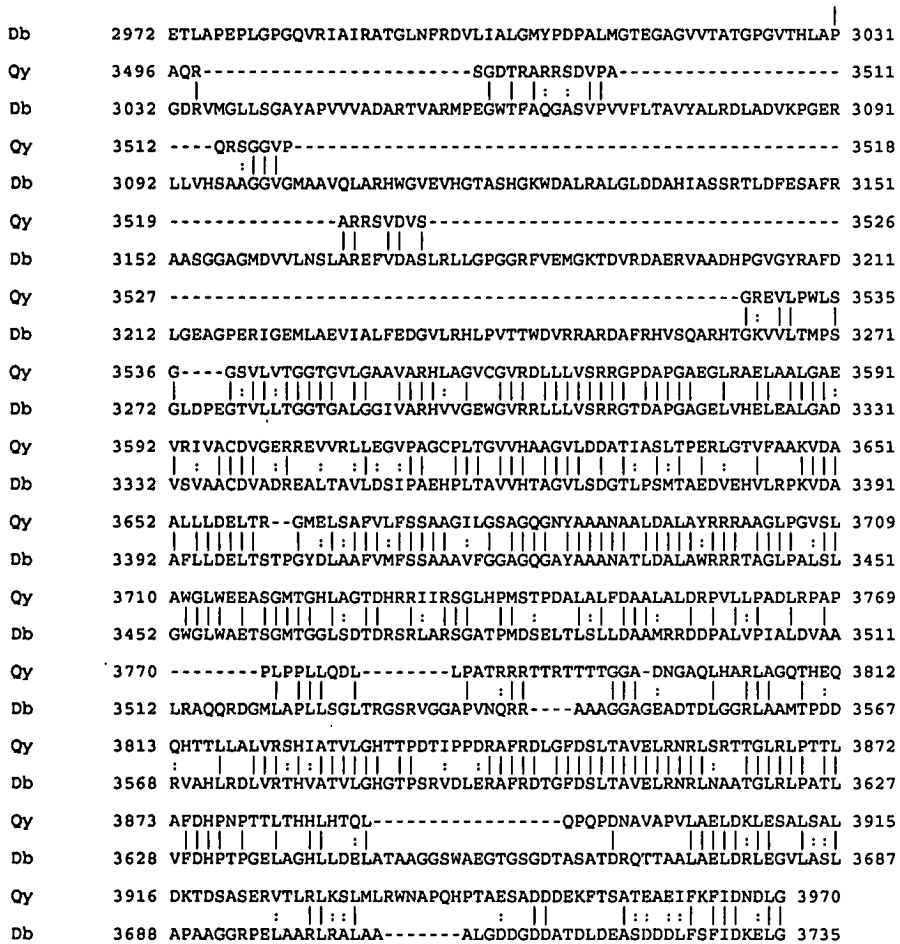
Qy       1301 FSDAVQALA-DDGHRVFVEVSPHPTLVPAIEDTTEDTAEDVTAIGSLRRGDNDTRRFLTA 1359
Db        821 FAPAVETLATDEGFTHFIEVSAHPVLTMTLPET-------VTGLGTLRREQGGQERLVTS 873

Qy       1360 LAHTHTTGIG------TPTTWHHHYTHHHTHPHPHTHLDLPTYPFQHQHYWLESSQPGAG 1413
Db        874 LAEAWTNGLTIDWAPVLPTATGH------------HPELPTYAFQRRHYWLHDSPAVQG 920

Qy       1414 SGSGAGAGSGAGSGRAGTAGGTAEVESRFWDAVARQDLETVATTLAVPPSAGLDTVVPAL 1473
Db        921 SVQ---------------------------------------------------------- 923
```


Qy 1474 SAWHRHQHDQARINTWTYQETWKPLTLPTTHQP---HQTWLIAIPETQTHHPHITNILTN 1530
Db 924 ----------DSWRYRIDWKRLAVADASERAGLSGRWLVVVPEDRS--AEAAPVLAA 968
Qy 1531 LHHHGITPIPLTLNHTHTNPQHLHHTLHHTRQQAQNHTTGAITGLLSLLALDETPHPHHP 1590
Db 969 LSGAGADPVQLDV----SPLGDRQRLAATLGEALAAAGGAVDGVLSLLAWDESAHPGHP 1023
Qy 1591 HTPT-GTLLNLTLTQTHTQTHPPTPLWYATTNATTTHPNDPLTHPTQAQTWGLARTTLLE 1649
Db 1024 APFTRGTGATLTLVQALEDAGVAAPLWCVTHGAVSVGRADHVTSPAQAMVWGMGRVAALE 1083
Qy 1650 HPTHTAGIIDLPTTPTPHTLQHLTQTLTQPHHQTQLAIRTTGTHTRRLTPTTLTPTHQPP 1709
Db 1084 HPERWGGLIDLPSDADRAALDRMTTVLAGGTGEDQVAVRASGLLARRLVRASL-PAHGTA 1142
Qy 1710 TP--TPHGTTLITGGTGALATHLTHHLTTHQPTQHLLLTSRTGPH----TPHAQH----- 1758
Db 1143 SPWWQADGTVLVTGAEEPAAAEAARRLARDGAGHLLLHTTPSGSEGAEGTSGAAEDSGLA 1202
Qy 1759 -LTTQLQQKGIHLTITTCDTSNPDQLQQLLNTIPPQHPLTTVIHTAGILDDATLTNLTPT 1817
Db 1203 GLVAELADLGATATVVTCDLTDAEAAARLLAGVSDAHPLSAVLHLPPTVDSEPLAATDAD 1262
Qy 1818 QLNNVLRAKAHSAHLLHQLTQHTPLT-----AFVLYSSAAATFGAPGQANYAAANAYLDA 1872
Db 1263 ALARVVTAKATAALHLDRLLREAAAAGGRPPVLVLFSSVAAIWGGAGQGAYAAGTAFLDA 1322
Qy 1873 LAHHRHTHHLPATSIAWGTWQGNGLADSDKARAYLDRRGFRPMSPELATAAVTQAI--AD 1930
Db 1323 LAGQHRADGPTVTSVAWSPWEGSRVTEGATGER-LRRLGLRPLAPATALTALDTALGHGD 1381
Qy 1931 TERPYVVIADIDWSKIE---HTSQTSDLVSAARAREPAV--QRPTPPAE---LHKTLAHQ 1982
Db 1382 TA---VTIADVDWSSFAPGFTTARPGTLLADLPEARRALDEQQSTTAADDTVLSRELGAL 1438
Qy 1983 TSADQRAALLELVRDHVAAVLRHADPKAIAPDQSFRALGFDSLTAVEFRNLLIKATGLRL 2042
Db 1439 TGAEQQRRMQELVREHLAVVLNHPSPEAVDTGRAFRDLGFDSLTAVELRNRLKNATGLAL 1498
Qy 2043 PVSLVFDHPTPAKLAVHLQNQLRGTAAESAPSAAAVTAEASV-TEPIAIVGMACRFPGGV 2101
Db 1499 PATLVFDYPTPRTLAEFLLAEILG---EQAGAGEQLPVDGGVDDEPVAIVGMACRLPGGV 1555
Qy 2102 TSADDFWDLISSEQDAIGGFPTDRGWDLDTLYDPDPDHPGTCYTRNGGFLYDAGHFDAEF 2161
Db 1556 ASPEDLWRLVAGGEDAISGFPQDRGWDVEGLYDPDPDASGRTYCRAGGFLDEAGEFDADF 1615
Qy 2162 FGISPREALAMDPQQRLLLETAWETIEHAGINPHTLHGTPTGVFTGTNGQDYALRVHNAG 2221
Db 1616 FGISPREALAMDPQQRLLLETSWEAVEDAGIDPTSLQGQQVGVFAGTNGPHYEPLLRNTA 1675
Qy 2222 QSTDGFALTGTAGSVISGRISYTFGFEGPAVSVDTACSSSLVALHLACQALRAGECSMAL 2281
Db 1676 EDLEGYVGTGNAASIMSGRVSYTLGLEGPAVTVDTACSSSLVALHLAVQALRKGECGLAL 1735
Qy 2282 AGGVTVMSSPGAFVEFSRQRGLAADGHCKAFSAAADGTGWGEGVGMLLVERLSDAHRNGH 2341
Db 1736 AGGVTVMSTPTTFVEFSRQRGLAEDGRSKAFAASADGFGPAEGVGMLLVERLSDARRNGH 1795
Qy 2342 RVLAVVRGSAVNQDGASNGLTAPNGPSQQRVIRQALANAGLSAGDVDAVEAHGTGTTLGD 2401
Db 1796 RVLAVVRGSAVNQDGASNGLTAPNGPSQQRVIRRALADARLTTADVDVVEAHGTGTRLGD 1855
Qy 2402 PIEAQALLATYGQDRAGEGPLWLGSVKSNVGHTQAAAGVAGVIKMVMALRHGLLPRTLHV 2461
Db 1856 PIEAQALIATYGQGRDTEQPLRLGSLKSNIGHTQAAAGVSGIIKMVQAMRHGVLPKTLHV 1915
Qy 2462 DEPSPHVDWSAGAVQLLTETVPWP-GGEGRLRRAGVSSFGVSGTNAHVILEEAPAD-DVP 2519
Db 1916 DRPSDQIDWSAGTVELLTEAMDWPRKQEGGLRRAAVSSFGISGTNAHIVLEEAPVDEDAP 1975

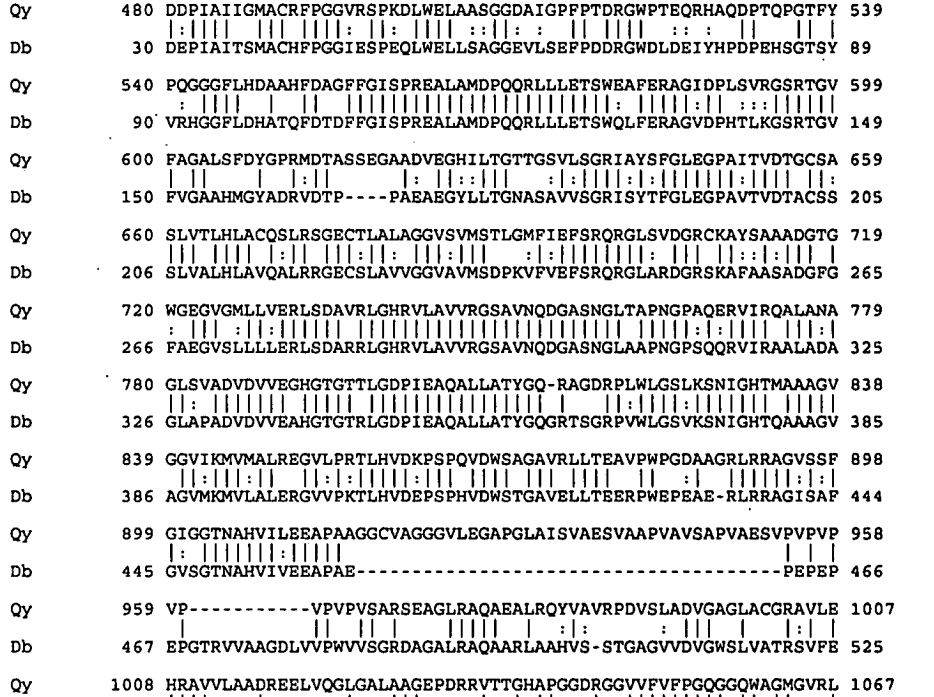

Qy 2520 GGPPAGEGDAGSDDEAAAGSPGVWPWLVSAKSQPALRAQAQALHAHLTDHPGLDLAD--- 2576
Db 1976 A----------DEPSVG--GVVPWLVSAKTPAALDAQIGRL-AAFASQGRTDAADPGA 2020
Qy 2577 VGYTLAHARAVFDHRATLIAADRDTFLQALQALAAGEPHPAVIHSSAPGG--TGTGEAAG 2634
Db 2021 VARVLAGGRAQFEHRAVALGTGQDDLAAAL--------------AAPEGLVRGVASGVG 2065
Qy 2635 KTAFICSGQGTQRPGMAHGLYHTHPVFAAALNDICTHLDPHLDHPLLPLLTQNDNDNEDA 2694
Db 2066 RVAFVFPGQGTQWAGMGAELLDVSKEFAAAMAECEAALAPYVDWSLEAVVRQ-----APG 2120
Qy 2695 AALLQQTRYAQPALFAFQVALHRLLTDGYHITPHYYAGHSLGEITAAHLAGILTLTDATT 2754
Db 2121 APTLERVDVVQPVTFAVMVSLAK-VWQHHGVTPQAVVGHSQGEIAAAYVAGALSLDDAAR 2179
Qy 2755 LITQRATLM--QTMPPGTMTTLHTTPHHITHHLTAHENDLAIAAINTPTSLVISGTPHTV 2812
Db 2180 VVTLRSKSIGAHLAGQGGMLSLALSEAAVVERLAGFDG-LSVAAVNGPTATVVSGDPTQI 2238
Qy 2813 QHITTLCQQQGIKTKTLPTNHAFHSPHTNPILNQLHQHTQTLTYHPPHTPLITA-----N 2867
Db 2239 QELAQACEADGVRARIIPVDYASHSAHVETIESELADVLAGLSPQTPQVPFFSTLEGAWI 2298
Qy 2868 TPPDQLLTPHYWTQQARNTVDYATTTQTL-HQHGVTTYIELGPDNTLTTLTHHNLPNPPT 2926
Db 2299 TEP--ALDGGYWYRNLRHRVGFAPAVETLATDEGFTHFVEVSAHPVLTMA----LPETVT 2352
Qy 2927 TTLTLTHPHHHPQTHLLTNLAKT-----TTTWHPHHYTHHDNQPHTHTHLDLPTYPFQHH 2981
Db 2353 GLGTLRR-DNGGQHRLTTSLAEAWANGLTVDW-------ASLLPTTTTHPDLPTYAFQTE 2404
Qy 2982 HYWLESTQPGAGNVSAAGLDPTEHPLLGATLELATDGGALLAGRLSLRSHPWLADHAVGG 3041
Db 2405 RYWPQPDLSAAGDITSAGLGAAEHPLLGAAVALADSDGCLLTGSLSLRTHPWLADHAVAG 2464
Qy 3042 TVLLSGATFLELALHAGTYVGCDRVDELTLHAPLVVPVDGGVSVQVGVAAADGEGRRLVS 3101
Db 2465 TVLLPGTAFVELAFRAGDQVGCDLVEELTLDAPLVLPRRGAVRVQLSVGASDESGRRTFG 2524
Qy 3102 VYARGGSACGGGGASGGVWTCHASGVLVEAAAGGVVVDGLAGV-----WPPRGAVAVDVD 3156
Db 2525 LYAHPEDAPG-----EAEWTRHATGVLAARA------DRTAPVADPEAWPPPGAEPVDVD 2573
Qy 3157 GVRDRLAGAGCVLGPVFSGLRAVWRDGGDLLAEVCLPEEAWG-DAAGFGLHPALLDGVVQ 3215
Db 2574 GLYERFAANGYGYGPLFQGVRGVWRRGDEVFADVALPAEVAGAEGARFGLHPALLDAAVQ 2633
Qy 3216 PLSVLLPGGTGFGEGAGFGEGVRVPAVWGGVSLHRAGVTGVRVRVSAVGRGGGREAVSVV 3275
Db 2634 ----------AAGAGGAFGAGTRLPFAWSGISLYAVGATALRVRLAP----AGPDTVSVS 2679
Qy 3276 VGDEAGVPVASVDRLELRPVDMGQLRAVSVSAGRRGSLYAVQW--------AEVGPVPVC 3327
Db 2680 AADSSGQPVFAADSLTVLPVDPAQLAAFSDPT--LDALHLLEWTAWDGAAQALPGAVVLG 2737
Qy 3328 GQA----------------WAWHEDVGES---GGGPVPGVVVLRCPDAGAGGGGGGGGGG 3368
Db 2738 GDADGLAAALRAGGTEVLSFPDLTDLVEAVDRGETPAPATVLVACPAAGP------GGPE 2791
Qy 3369 GVGEVVGGVLGVVQGWLGLERFAGSRLVVVTRGAVVAGPEDGPVDVVGASVWGLVRSAQA 3428
Db 2792 HVREALHGSLALMQAWLADERFTDGRLVLVTRDAVAARSGDGLRSTGQAAVWGLGRSAQT 2851
Qy 3429 EHPDRFVLLDL--------DTDTGTDLDTGAGAGWGVD-----GGRVAAVVACGEPQLAV 3475
Db 2852 ESPGRFVLLDLAGEARTAGDATAGDGLTTGDATVGGTSGDAALGSALATALGSGEPQLAL 2911
Qy 3476 RGERLLAARLKR----------LES----SGD- 3493
Db 2912 RDGALLVPRLARAAAPAAADGLAAADGLAALPLPAAPALWRLEPGTDGSLESLTAAPGDA 2971
Qy 3494 ----------VP 3495

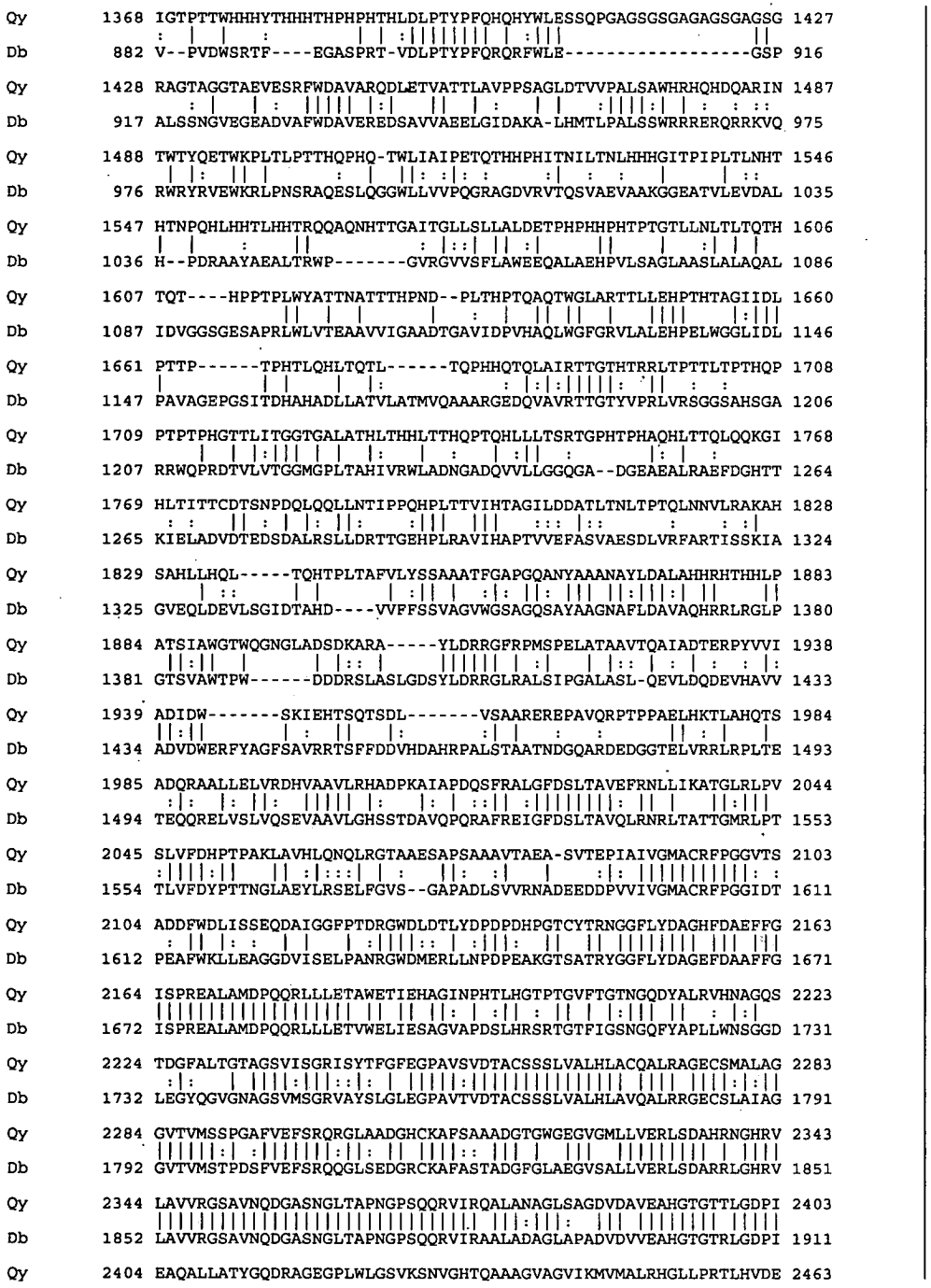


RESULT 11
US-09-808-880-3
; Sequence 3, Application US/09808880
; Publication No. US20030027287A1
; GENERAL INFORMATION:
; APPLICANT: Betlach, Mary C.
; APPLICANT: Shah, Sanjay Krishnakant
; APPLICANT: McDaniel, Robert
; APPLICANT: Tang, Li
; TITLE OF INVENTION: RECOMBINANT OLEANDOLIDE POLYKETIDE SYNTHASE
; FILE REFERENCE: 30062-20029.00
; CURRENT APPLICATION NUMBER: US/09/808,880
; CURRENT FILING DATE: 2001-03-14
; PRIOR APPLICATION NUMBER: US/09/428,517
; PRIOR FILING DATE: 1999-10-28
; PRIOR APPLICATION NUMBER: 60/120,254
; PRIOR FILING DATE: 1999-02-16
; PRIOR APPLICATION NUMBER: 60/106,100
; PRIOR FILING DATE: 1998-10-29
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 3
; LENGTH: 3816
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Recombinant
; OTHER INFORMATION: Oleandolide PKS
US-09-808-880-3

Query Match 33.8%; Score 7041.5; DB 9; Length 3816;
Best Local Similarity 41.7%; Pred. No. 0;
Matches 1654; Conservative 450; Mismatches 1206; Indels 655; Gaps 82;





```
Db       2928 TASRLVLVTRHAMTVAESDPEPDLLLAPVWGLVRSAQAENPGRFVLADIDGDEASWDALP 2987

Qy       3441 -------------------------TDTGTDLDTGAGAGWGVD---GGRVA--AVVAC-- 3468

Db       2988 RAVASAASEVAIRAGAVYVPRLARATDEGLVVADEAAGPWRLDVTEAGTLANLALVPCPD 3047

Qy       3469 -------GEPQLAVRG-----------------ERLLAA--------------------R 3484

Db       3048 ASRPLGPDEVRIAVRAAGVNFRDVLLALGMYPDEGLMGAEAAGVVTEVGGGVTTLAPGDR 3107

Qy       3485 LKRLESSG---------------------------------------DVPAQRSG----- 3500

Db       3108 VMGLVTGGFGPVAVTHHRMLVRMPRGWSFAEAASVPVAFLTAYYALHDLAGLRGGESVLV 3167

Qy       3501 ------------------------------------------------------------ 3500

Db       3168 HSAAGGVGMAAVQLARHWDAEVFGTASKGKWDVLAAQGLDEEHIGSSRTTEFEQRFRATS 3227

Qy       3501 -----------------DTRAR---------------RSDVPAQRSGGVPARRSV--DV 3525

Db       3228 GGRGIDVVLNALSGDFVDASARLLREGGRFVEMGKTDIRTDLGVVGADGVPDIRYVAFDL 3287

Qy       3526 S---------------------------------GREVLPWLS--------------- 3535

Db       3288 AEAGAERIGQMLDEIMALFDAGVLRLPPLRAWPVRRAHEALRFVSQARHVGKVVLTVPAA 3347

Qy       3536 ---GGSVLVTGGTGVLGAAVARHLAGVCGVRDLLLVSRRG--PDAPGAEGLRAELAALGA 3590

Db       3348 LDAEGTVLIT-GAGTLGALVARHLVTEHDVRRLLLVSRSGVAPD------LAAELGALGA 3400

Qy       3591 EVRIVACDVGERREVVRLLEGVPAGCPLTGVVHAAGVLDDATIASLTPERLGTVFAAKVD 3650

Db       3401 EVTVAACDVANRKALKALLEDIPPEHPVTGIVHTAGVLDDGVVSGLTPERVDTVLKPKVD 3460

Qy       3651 AALLLDELTRGMEL--SAFVLFSSAAGILGSAGQGNYAAANAALDALAYRRRAAGLPGVS 3708

Db       3461 AALTLESVIGELDLDPALFVIFSSAASMLGGPGQGSYAAANQFLDTLARHRARRGLTSVS 3520

Qy       3709 LAWGLWEEASGMTGHLAGTDHRRIIRSGLHPMSTPDALALFDAALALDRPVLLPADLRPA 3768

Db       3521 LGWGLWHEASGLTGGLADIDRDRMSRAGIAPMPTDEALHLFDRATELGDPVLLPMRLNEA 3580

Qy       3769 P--------PLPPLLQDLLPATRRRTTRT-TTTGGADNGAQLHARLAGQTHEQQHTTLLA 3819

Db       3581 ALEDRAADGTLPPLLSGLVRVRHRPSARAGTATAAPATGPEAFARELAAAPDPRR-ALRD 3639

Qy       3820 LVRSHIATVLGHTTPDTIPPDRAFRDLGFDSLTAVELRNRLSRTTGLRLPTTLAFDHPNP 3879

Db       3640 LVRGHVALVLGHSGPEAIDAEQAFRDIGFDSLTAVELRNRLNAETGLRLPGTLVFDYPNP 3699

Qy       3880 TTLTHHLHTQLQPQPDNAVAPVLAELDKLESALSALDKTDSASERV--------TLRLKS 3931

Db       3700 SALADHLLELLAPATQPTAAPLLAELERVEQLLSAAASPGGPASAVDEETRTLIATRLAT 3759

Qy       3932 LMLRW---------NAP-----------------QHPTAESADDDEKFTSATEAEIFKF 3964

Db       3760 LASQWTHLPVGSPGNADNRSGPGESGQAQESGATGEHTAAWTSDDD----------LFAF 3809

Qy       3965 IDNDL 3969

Db       3810 LDKRL 3814


RESULT 12
US-09-860-846-33
; Sequence 33, Application US/09860846
; Patent No. US20020164742A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/860,846
;  CURRENT FILING DATE:  2001-05-18
;  PRIOR APPLICATION NUMBER: 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 33
;   LENGTH: 3739
;   TYPE: PRT
;   ORGANISM: Streptomyces venezuelae
US-09-860-846-33

  Query Match             33.7%;  Score 7017;  DB 9;  Length 3739;
  Best Local Similarity   42.4%;  Pred. No. 0;
  Matches 1676;  Conservative  390;  Mismatches 1177;  Indels  712;  Gaps   82;

Qy        477 AAADDPIAIIGMACRFPGGVRSPKDLWELAASGGDAIGPFPTDRGWPTEQRHAQDPTQPG 536

Db         32 AKAGEPVAIVGMACRLPGGVASPEDLWRLVAGGEDAISEFPQDRGWDVEGLYDPNPEATG 91

Qy        537 TFYPQGGGFLHDAAHFDAGFFGISPREALAMDPQQRLLLETSWEAFERAGIDPLSVRGSR 596

Db         92 KSYAREAGFLYEAGEFDADFFGISPREALAMDPQQRLLLEASWEAFEHAGIPAATARGTS 151

Qy        597 TGVFAGALSFDYGPRMDTASSEGAADVEGHILTGTTGSVLSGRIAYSFGLEGPAITVDTG 656

Db        152 VGVFTGVMYHDYATRL-TDVPEG---IEGYLGTGNSGSVASGRVAYTLGLEGPAVTVDTA 207

Qy        657 CSASLVTLHLACQSLRSGECTLALAGGVSVMSTLGMFIEFSRQRGLSVDGRCKAYSAAAD 716

Db        208 CSSSLVALHLAVQALRKGEVDMALAGGVTVMSTPSTFVEFSRQRGLAPDGRSKSFSSTAD 267

Qy        717 GTGWGEGVGMLLVERLSDAVRLGHRVLAVVRGSAVNQDGASNGLTAPNGPAQERVIRQAL 776

Db        268 GTSWSEGVGVLLVERLSDARRKGHRILAVVRGTAVNQDGASSGLTAPNGPSQQRVIRRAL 327

Qy        777 ANAGLSVADVDVVEGHGTGTTLGDPIEAQALLATYGQ-RAGDRPLWLGSLKSNIGHTMAA 835

Db        328 ADARLTTSDVDVVEAHGTGTRLGDPIEAQAVIATYGQGRDGEQPLRLGSLKSNIGHTQAA 387

Qy        836 AGVGGVIKMVMALREGVLPRTLHVDKPSPQVDWSAGAVRLLTEAVPWPGDAAGRLRRAGV 895

Db        388 AGVSGVIKMVQAMRHGVLPKTLHVEKPTDQVDWSAGAVELLTEAMDWPDKGDGGLRRAAV 447

Qy        896 SSFGIGGTNAHVILEEAPAAGGCVAGGGVLEGAPGLAISVAESVAAPVAVSAPVAESVPV 955

Db        448 SSFGVSGTNAHVVLEEAPA----------------------------AEETPASEATPA 478

Qy        956 PVPVP----VPVPVSARSEAGLRAQAEALRQYVAVRPDVSLADVGA---GLACGRAVLEH 1008

Db        479 VEPSVGAGLVPWLVSAKTPAALDAQIGRLAAF-ASQGRTDAADPGAVARVLAGGRAEFEH 537

Qy       1009 RAVVLAADREELVQGLGA---LAAGEPDRRVTTGHAPGGDRGGVVFVFPGQGGQWAGMGV 1065

Db        538 RAVVLGTGQDDFAQALTAPEGLIRGTP----------SDVGRVAFVFPGQGTQWAGMGA 586

Qy       1066 RLLASSPVFARRMQACEEALAPWVDWSVVDILRRDAGDAVWERADVVQPVLFSVMVSLAA 1125

Db        587 ELLDVSKEFAAAMAECESALSRYVDWSLEAVVRQAPGAPTLERVDVVQPVTFAVMVSLAK 646

Qy       1126 LWRSYGIEPDAVLGHSQGEIAAAHVCGALSLKDAAKTVALRSRALAA-VRGRGGMASVPL 1184

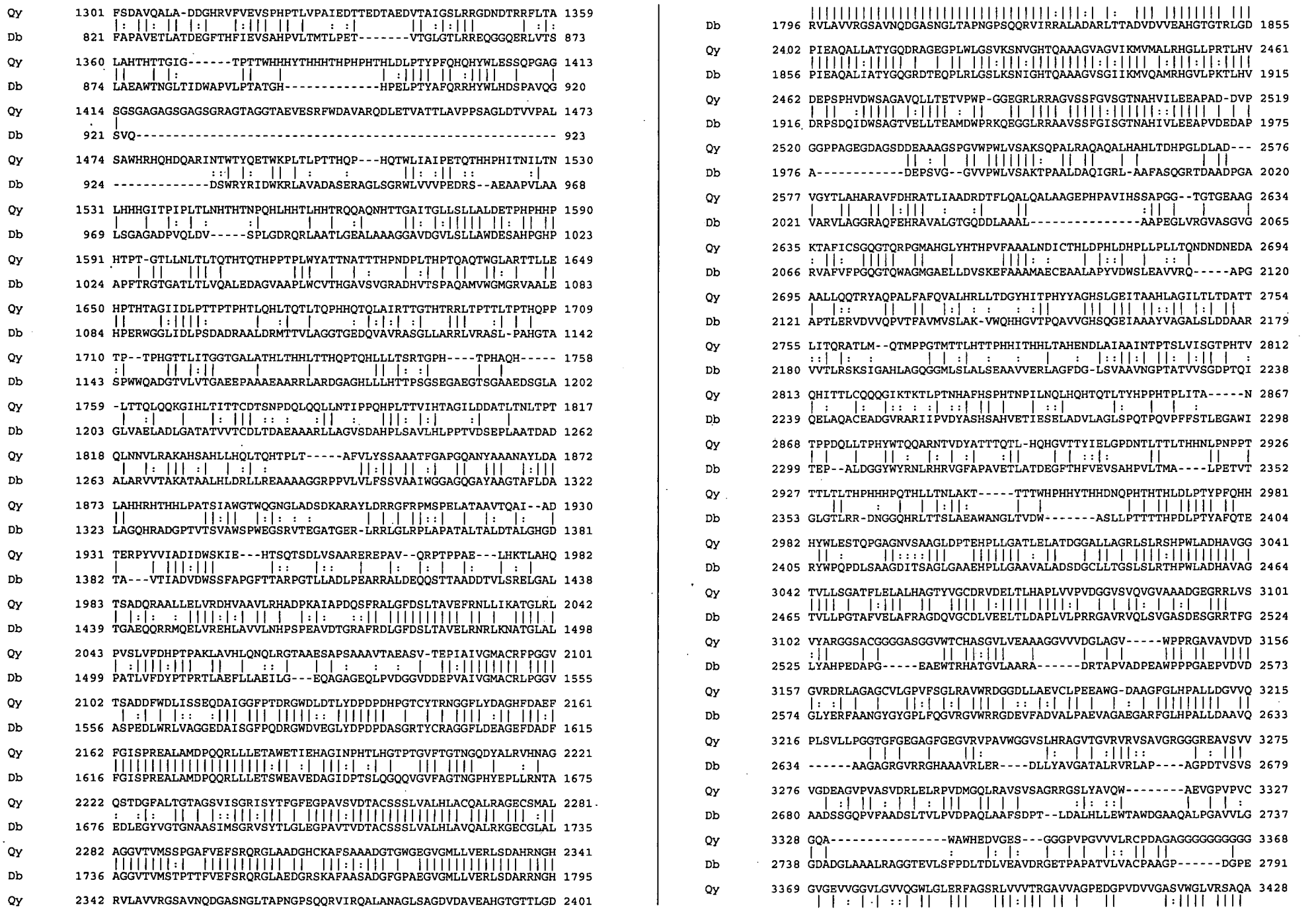
Db        647 VWQHHGVTPQAVVGHSQGEIAAAYVAGALTLDDAARVVTLRSKSIAAHLAGKGGMISLAL 706

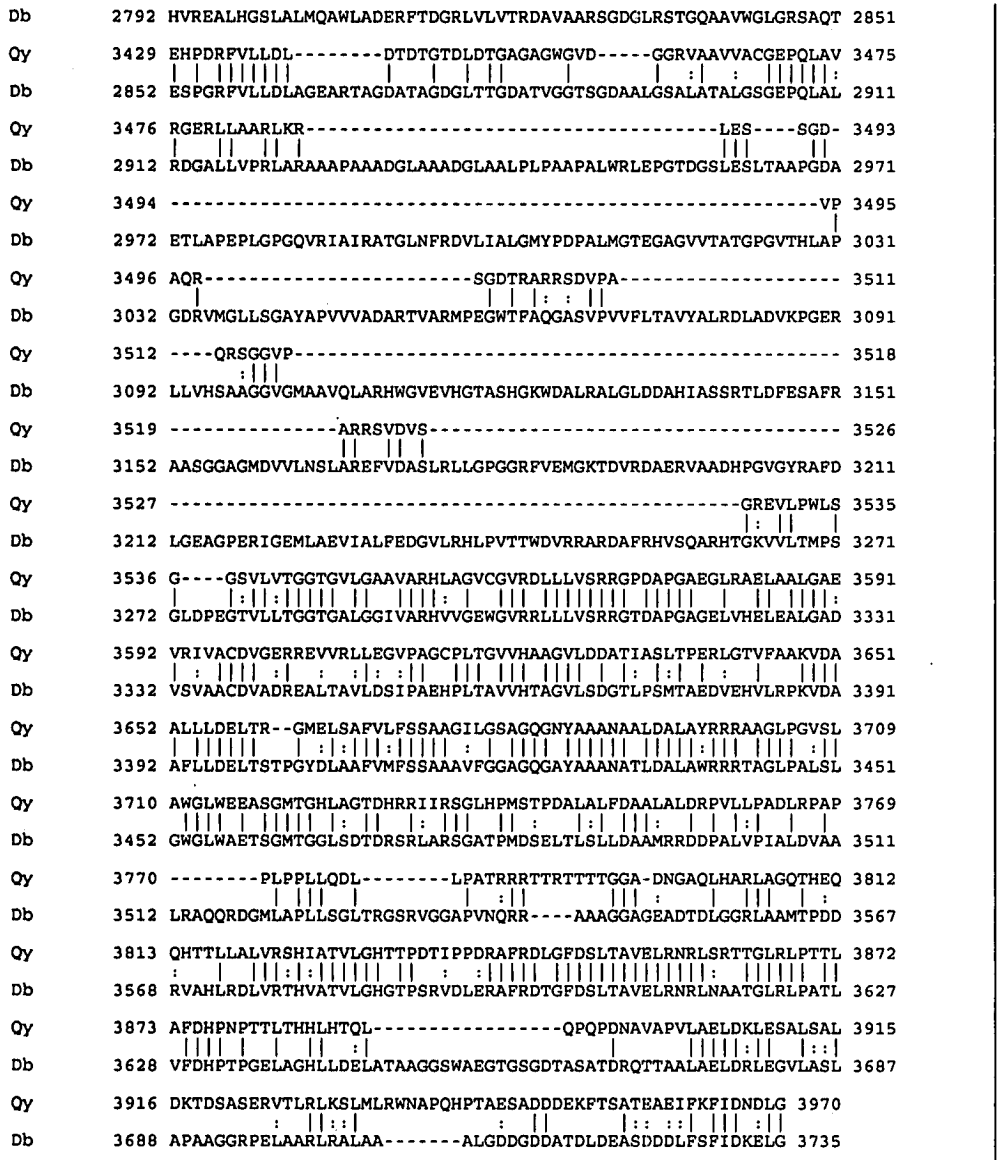
Qy       1185 ----PAQEVEQLIGERWAGRLWVAAVNGPRSTAVSGDAEAVDEVLAYCAGTGVRARRIPV 1240

Db        707 SEEATRQRIENLHG------LSIAAVNGPTATVVSGDPTQIQELAQACEADGVRARIIPV 760

Qy       1241 DYASHCPHVQPLREELLELLGDISPQPSGVPFFSTVEGTWLDTTTLDAAYWYRNLHQPVR 1300

Db        761 DYASHSAHVETIESELAEVLAGLSPRTPEVPFFSTLEGAWITEPVLDGTYWYRNLRHRVG 820
```

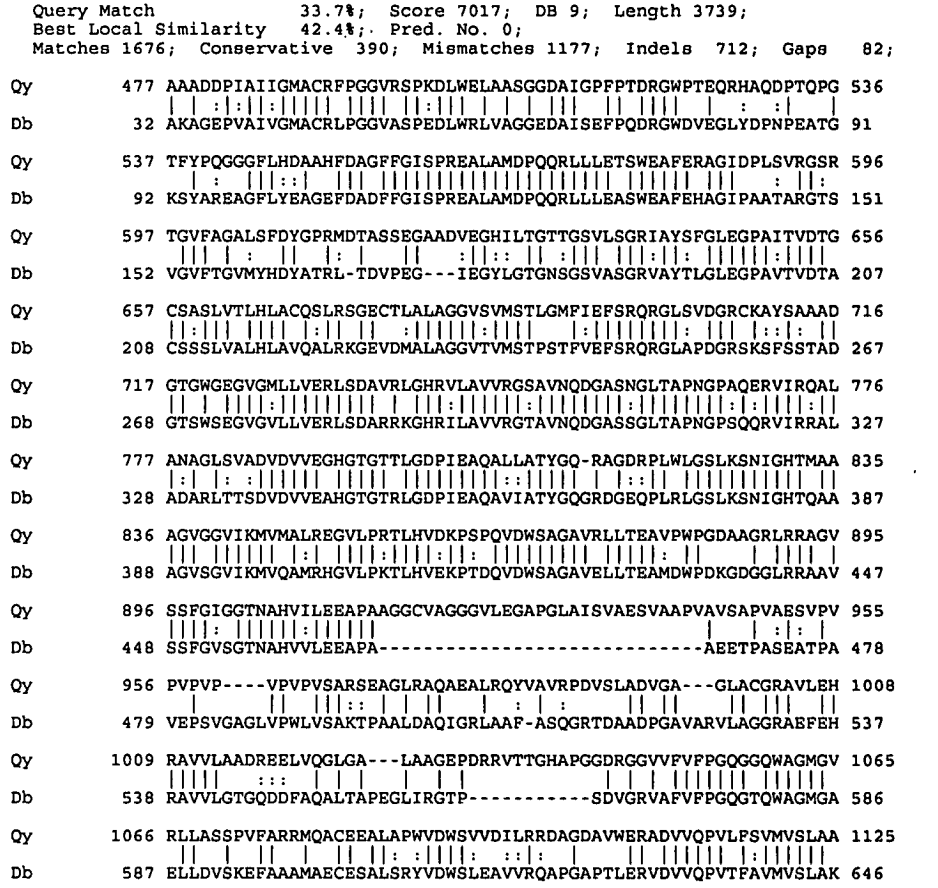


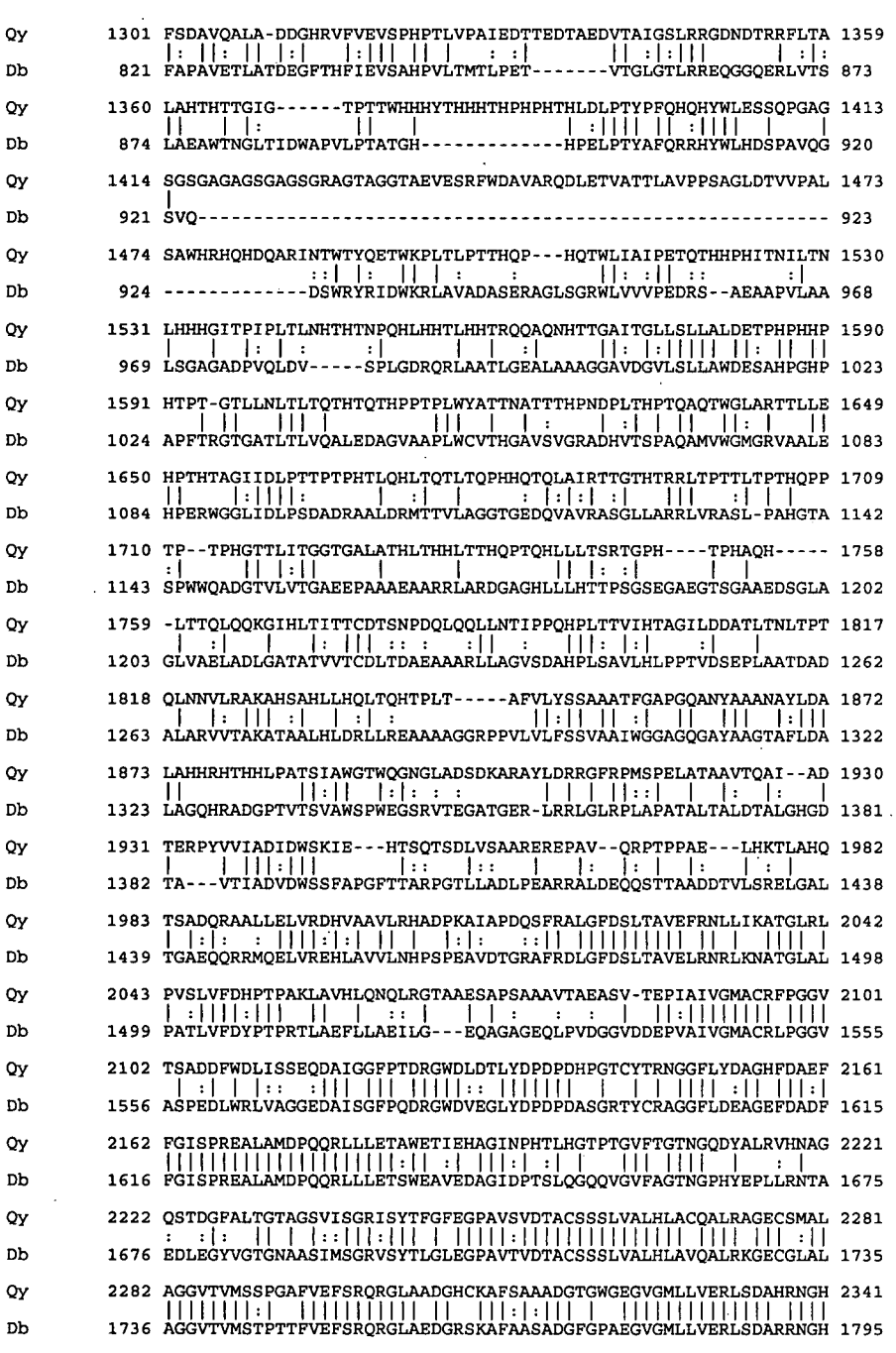


RESULT 13
US-09-988-384B-33
; Sequence 33, Application US/09988384B
; Publication No. US20030073824A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin

; FILE REFERENCE: 600.536US1
; CURRENT APPLICATION NUMBER: US/09/988,384B
; CURRENT FILING DATE: 2001-11-19
; PRIOR APPLICATION NUMBER: PCT/US99/14398
; PRIOR FILING DATE: 1999-06-25
; PRIOR APPLICATION NUMBER: US 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 53
; SEQ ID NO 33
; LENGTH: 3739
; TYPE: PRT
; ORGANISM: Streptomyces venezuelae
US-09-988-384B-33

Query Match 33.7%; Score 7017; DB 9; Length 3739;
Best Local Similarity 42.4%; Pred. No. 0;
Matches 1676; Conservative 390; Mismatches 1177; Indels 712; Gaps 82;




Qy 1301 FSDAVQALA-DDGHRVFVEVSPHPTLVPAIEDTTEDTAEDVTAIGSLRRGDNDTRRFLTA 1359
Db 821 FAPAVETLATDEGFTHFIEVSAHPVLTMTLPET-------VTGLGTLRREQGGQERLVTS 873
Qy 1360 LAHTHTTGIG------TPTTWHHHYTHHHTHPHPHTHLDLPTYPFQHQHYWLESSQPGAG 1413
Db 874 LAEAWTNGLTIDWAPVLPTATGH-------------HPELPTYAFQRRHYWLHDSPAVQG 920
Qy 1414 SGSGAGAGSGAGSGRAGTAGGTAEVESRFWDAVARQDLETVATTLAVPPSAGLDTVVPAL 1473
Db 921 SVQ---------- 923
Qy 1474 SAWHRHQHDQARINTWTYQETWKPLTLPTTHQP---HQTWLIAIPETQTHHPHITNILTN 1530
Db 924 ------------DSWRYRIDWKRLAVADASERAGLSGRWLVVVPEDRS--AEAAPVLAA 968
Qy 1531 LHHHGITPIPLTLNHTHTNPQHLHHTLHHTRQQAQNHTTGAITGLLSLLALDETPHPHHP 1590
Db 969 LSGAGADPVQLDV-----SPLGDRQRLAATLGEALAAAGGAVDGVLSLLAWDESAHPGHP 1023
Qy 1591 HTPT-GTLLNLTLTQTHTQTHPPTPLWYATTNATTTHPNDPLTHPTQAQTWGLARTTLLE 1649
Db 1024 APFTRGTGATLTLVQALEDAGVAAPLWCVTHGAVSVGRADHVTSPAQAMVWGMGRVAALE 1083
Qy 1650 HPTHTAGIIDLPTTPTPHTLQHLTQTLTQPHHQTQLAIRTTGTHTRRLTPTTLTPTHQPP 1709
Db 1084 HPERWGGLIDLPSDADRAALDRMTTVLAGGTGEDQVAVRASGLLARRLVRASL-PAHGTA 1142
Qy 1710 TP--TPHGTTLITGGTGALATHLTHHLTTHQPTQHLLLTSRTGPH----TPHAQH----- 1758
Db 1143 SPWWQADGTVLVTGAEEPAAAEAARRLARDGAGHLLLHTTPSGSEGAEGTSGAAEDSGLA 1202
Qy 1759 -LTTQLQQKGIHLTITTCDTSNPDQLQQLLNTIPPQHPLTTVIHTAGILDDATLTNLTPT 1817
Db 1203 GLVAELADLGATATVVTCDLTDAEAAARLLAGVSDAHPLSAVLHLPPTVDSEPLAATDAD 1262
Qy 1818 QLNNVLRAKAHSAHLLHQLTQHTPLT-----AFVLYSSAAATFGAPGQANYAAANAYLDA 1872
Db 1263 ALARVVTAKATAALHLDRLLREAAAAGGRPPVLVLFSSVAAIWGGAGQGAYAAGTAFLDA 1322
Qy 1873 LAHHRHTHHLPATSIAWGTWQGNGLADSDKARAYLDRRGFRPMSPELATAAVTQAI--AD 1930
Db 1323 LAGQHRADGPTVTSVAWSPWEGSRVTEGATGER-LRRLGLRPLAPATALTALDTALGHGD 1381
Qy 1931 TERPYVVIADIDWSKIE---HTSQTSDLVSAARAREPAV--QRPTPPAE---LHKTLAHQ 1982
Db 1382 TA---VTIADVDWSSFAPGFTTARPGTLLADLPEARRALDEQQSTTAADDTVLSRELGAL 1438
Qy 1983 TSADQRAALLELVRDHVAAVLRHADPKAIAPDQSFRALGFDSLTAVEFRNLLIKATGLRL 2042
Db 1439 TGAEQQRRMQELVREHLAVVLNHPSPEAVDTGRAFRDLGFDSLTAVELRNRLKNATGLAL 1498
Qy 2043 PVSLVFDHPTPAKLAVHLQNQLRGTAAESAPSAAAVTAEASV-TEPIAIVGMACRFPGGV 2101
Db 1499 PATLVFDYPTPRTLAEFLLAEILG---EQAGAGEQLPVDGGVDDEPVAIVGMACRLPGGV 1555
Qy 2102 TSADDFWDLISSEQDAIGGFPTDRGWDLDTLYDPDPDHPGTCYTRNGGFLYDAGHFDAEF 2161
Db 1556 ASPEDLWRLVAGGEDAISGFPQDRGWDVEGLYDPDPDASGRTYCRAGGFLDEAGEFDADF 1615
Qy 2162 FGISPREALAMDPQQRLLLETAWETIEHAGINPHTLHGTPTGVFTGTNGQDYALRVHNAG 2221
Db 1616 FGISPREALAMDPQQRLLLETSWEAVEDAGIDPTSLQGQQVGVFAGTNGPHYEPLLRNTA 1675
Qy 2222 QSTDGFALTGTAGSVISGRISYTFGFEGPAVSVDTACSSSLVALHLACQALRAGECSMAL 2281
Db 1676 EDLEGYVGTGNAASIMSGRVSYTLGLEGPAVTVDTACSSSLVALHLAVQALRKGECGLAL 1735
Qy 2282 AGGVTVMSSPGAFVEFSRQRGLAADGHCKAFSAAADGTGWGEGVGMLLVERLSDAHRNGH 2341
Db 1736 AGGVTVMSTPTTFVEFSRQRGLAEDGRSKAFAASADGFGPAEGVGMLLVERLSDARRNGH 1795
Qy 2342 RVLAVVRGSAVNQDGASNGLTAPNGPSQQRVIRQALANAGLSAGDVDAVEAHGTGTTLGD 2401
Db 1796 RVLAVVRGSAVNQDGASNGLTAPNGPSQQRVIRRALADARLTTADVDVVEAHGTGTRLGD 1855
Qy 2402 PIEAQALLATYGQDRAGEGPLWLGSVKSNVGHTQAAAGVAGVIKMVMALRHGLLPRTLHV 2461
Db 1856 PIEAQALIATYGQGRDTEQPLRLGSLKSNIGHTQAAAGVSGIIKMVQAMRHGVLPKTLHV 1915
Qy 2462 DEPSPHVDWSAGAVQLLTETVPWP-GGEGRLRRAGVSSFGVSGTNAHVILEEAPAD-DVP 2519
Db 1916 DRPSDQIDWSAGTVELLTEAMDWPRKQEGGLRRAAVSSFGISGTNAHIVLEEAPVDEDAP 1975
Qy 2520 GGPPAGEGDAGSDDEAAAGSPGVWPWLVSAKSQPALRAQAQALHAHLTDHPGLDLAD--- 2576
Db 1976 A------------DEPSVG--GVVPWLVSAKTPAALDAQIGRL-AAFASQGRTDAADPGA 2020
Qy 2577 VGYTLAHARAVFDHRATLIAADRDTFLQALQALAAGEPHPAVIHSSAPGG--TGTGEAAG 2634
Db 2021 VARVLAGGRAQFEHRAVALGTGQDDLAAAL---------------AAPEGLVRGVASGVG 2065
Qy 2635 KTAFICSGQGTQRPGMAHGLYHTHPVFAAALNDICTHLDPHLDHPLLPLLTQNDNDNEDA 2694
Db 2066 RVAFVFPGQGTQWAGMGAELLDVSKEFAAAMAECEAALAPYVDWSLEAVVRQ-----APG 2120
Qy 2695 AALLQQTRYAQPALFAFQVALHRLLTDGYHITPHYYAGHSLGEITAAHLAGILTLTDATT 2754
Db 2121 APTLERVDVVQPVTFAVMVSLAK-VWQHHGVTPQAVVGHSQGEIAAAYVAGALSLDDAAR 2179
Qy 2755 LITQRATLM--QTMPPGTMTTLHTTPHHITHHLTAHENDLAIAAINTPTSLVISGTPHTV 2812
Db 2180 VVTLRSKSIGAHLAGQGGMLSLALSEAAVVERLAGFDG-LSVAAVNGPTATVVSGDPTQI 2238
Qy 2813 QHITTLCQQQGIKTKTLPTNHAFHSPHTNPILNQLHQHTQTLTYHPPHTPLITA-----N 2867
Db 2239 QELAQACEADGVRARIIPVDYASHSAHVETIESELADVLAGLSPQTPQVPFFSTLEGAWI 2298
Qy 2868 TPPDQLLTPHYWTQQARNTVDYATTTQTL-HQHGVTTYIELGPDNTLTTLTHHNLPNPPT 2926
Db 2299 TEP--ALDGGYWYRNLRHRVGFAPAVETLATDEGFTHFVEVSAHPVLTMA----LPETVT 2352
Qy 2927 TTLTLTHPHHHPQTHLLTNLAKT-----TTTWHPHHYTHHDNQPHTHTHLDLPTYPFQHH 2981
Db 2353 GLGTLRR-DNGGQHRLTTSLAEAWANGLTVDW-------ASLLPTTTTHPDLPTYAFQTE 2404
Qy 2982 HYWLESTQPGAGNVSAAGLDPTEHPLLGATLELATDGGALLAGRLSLRSHPWLADHAVGG 3041
Db 2405 RYWPQPDLSAAGDITSAGLGAAEHPLLGAAVALADSDGCLLTGSLSLRTHPWLADHAVAG 2464
Qy 3042 TVLLSGATFLELALHAGTYVGCDRVDELTLHAPLVVPVDGGVSVQVGVAAADGEGRRLVS 3101
Db 2465 TVLLPGTAFVELAFRAGDQVGCDLVEELTLDAPLVLPRRGAVRVQLSVGASDESGRRTFG 2524
Qy 3102 VYARGGSACGGGGASGGVWTCHASGVLVEAAAGGVVVDGLAGV-----WPPRGAVAVDVD 3156
Db 2525 LYAHPEDAPG-----EAEWTRHATGVLAARA------DRTAPVADPEAWPPPGAEPVDVD 2573
Qy 3157 GVRDRLAGAGCVLGPVFSGLRAVWRDGGDLLAEVCLPEEAWG-DAAGFGLHPALLDGVVQ 3215
Db 2574 GLYERFAANGYGYGPLFQGVRGVWRRGDEVFADVALPAEVAGAEGARFGLHPALLDAAVQ 2633
Qy 3216 PLSVLLPGGTGFGEGAGFGEGVRVPAVWGGVSLHRAGVTGVRVRVSAVGRGGGREAVSVV 3275
Db 2634 ------AAGAGRGVRRGHAAAVRLER----DLLYAVGATALRVRLAP----AGPDTVSVS 2679
Qy 3276 VGDEAGVPVASVDRLELRPVDMGQLRAVSVSAGRRGSLYAVQW--------AEVGPVPVC 3327
Db 2680 AADSSGQPVFAADSLTVLPVDPAQLAAFSDPT--LDALHLLEWTAWDGAAQALPGAVVLG 2737
Qy 3328 GQA----------------WAWHEDVGES---GGGPVPGVVVLRCPDAGAGGGGGGGGGG 3368
Db 2738 GDADGLAAALRAGGTEVLSFPDLTDLVEAVDRGETPAPATVLVACPAAGP------DGPE 2791
Qy 3369 GVGEVVGGVLGVVQGWLGLERFAGSRLVVVTRGAVVAGPEDGPVDVVGASVWGLVRSAQA 3428

```
Db       2792 HVREALHGSLALMQAWLADERFTDGRLVLVTRDAVAARSGDGLRSTGQAAVWGLGRSAQT 2851

Qy       3429 EHPDRPVLLDL--------DTDTGTDLDTGAGAGWGVD-----GGRVAAVVACGEPQLAV 3475

Db       2852 ESPGRPVLLDLAGEARTAGDATAGDGLTTGDATVGGTSGDAALGSALATALGSGEPQLAL 2911

Qy       3476 RGERLLAARLKR-----------------------------------LES----SGD- 3493

Db       2912 RDGALLVPRLARAAAPAAADGLAAADGLAALPLPAAPALWRLEPGTDGSLESLTAAPGDA 2971

Qy       3494 ------------------------------------------------------VP 3495

Db       2972 ETLAPEPLGPGQVRIAIRATGLNFRDVLIALGMYPDPALMGTEGAGVVTATGPGVTHLAP 3031

Qy       3496 AQR-----------------------SGDTRARRSDVPA-------------------- 3511

Db       3032 GDRVMGLLSGAYAPVVVADARTVARMPEGWTFAQGASVPVVFLTAVYALRDLADVKPGER 3091

Qy       3512 ----QRSGGVP------------------------------------------------- 3518

Db       3092 LLVHSAAGGVGMAAVQLARHWGVEVHGTASHGKWDALRALGLDDAHIASSRTLDFESAFR 3151

Qy       3519 ---------------ARRSVDVS----------------------------------- 3526

Db       3152 AASGGAGMDVVLNSLAREFVDASLRLLGPGGRFVEMGKTDVRDAERVAADHPGVGYRAFD 3211

Qy       3527 ------------------------------------------------GREVLPWLS 3535

Db       3212 LGEAGPERIGEMLAEVIALFEDGVLRHLPVTTWDVRRARDAFRHVSQARHTGKVVLTMPS 3271

Qy       3536 G----GSVLVTGGTGVLGAAVARHLAGVCGVRDLLLVSRRGPDAPGAEGLRAELAALGAE 3591

Db       3272 GLDPEGTVLLTGGTGALGGIVARHVVGEWGVRRLLLVSRRGTDAPGAGELVHELEALGAD 3331

Qy       3592 VRIVACDVGERREVVRLLEGVPAGCPLTGVVHAAGVLDDATIASLTPERLGTVFAAKVDA 3651

Db       3332 VSVAACDVADREALTAVLDSIPAEHPLTAVVHTAGVLSDGTLPSMTAEDVEHVLRPKVDA 3391

Qy       3652 ALLLDELTR--GMELSAFVLFSSAAGILGSAGQGNYAAANAALDALAYRRRAAGLPGVSL 3709

Db       3392 AFLLDELTSTPGYDLAAFVMFSSAAAVFGGAGQGAYAAANATLDALAWRRRTAGLPALSL 3451

Qy       3710 AWGLWEEASGMTGHLAGTDHRRIIRSGLHPMSTPDALALFDAALALDRPVLLPADLRPAP 3769

Db       3452 GWGLWAETSGMTGGLSDTDRSRLARSGATPMDSELTLSLLDAAMRRDDPALVPIALDVAA 3511

Qy       3770 --------PLPPLLQDL--------LPATRRRTTRTTTTGGA-DNGAQLHARLAGQTHEQ 3812

Db       3512 LRAQQRDGMLAPLLSGLTRGSRVGGAPVNQRR----AAAGGAGEADTDLGGRLAAMTPDD 3567

Qy       3813 QHTTLLALVRSHIATVLGHTTPDTIPPDRAFRDLGFDSLTAVELRNRLSRTTGLRLPTTL 3872

Db       3568 RVAHLRDLVRTHVATVLGHGTPSRVDLERAFRDTGFDSLTAVELRNRLNAATGLRLPATL 3627

Qy       3873 AFDHPNPTTLTHHLHTQL-----------------QPQPDNAVAPVLAELDKLESALSAL 3915

Db       3628 VFDHPTPGELAGHLLDELATAAGGSWAEGTGSGDTASATDRQTTAALAELDRLEGVLASL 3687

Qy       3916 DKTDSASERVTLRLKSLMLRWNAPQHPTAESADDDEKFTSATEAEIFKFIDNDLG 3970

Db       3688 APAAGGRPELAARLRALAA-------ALGDDGDDATDLDEASDDDLFSFIDKELG 3735


RESULT 14
US-09-836-821-33
; Sequence 33, Application US/09836821
; Publication No. US20030087405A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/836,821
;  CURRENT FILING DATE:  2001-04-17
;  PRIOR APPLICATION NUMBER: 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 33
;   LENGTH: 3739
;   TYPE: PRT
;   ORGANISM: Streptomyces venezuelae
US-09-836-821-33

  Query Match             33.7%;  Score 7017;  DB 9;  Length 3739;
  Best Local Similarity   42.4%;  Pred. No. 0;
  Matches 1676;  Conservative  390;  Mismatches 1177;  Indels  712;  Gaps   82;

Qy        477 AAADDPIAIIGMACRFPGGVRSPKDLWELAASGGDAIGPFPTDRGWPTEQRHAQDPTQPG 536

Db         32 AKAGEPVAIVGMACRLPGGVASPEDLWRLVAGGEDAISEFPQDRGWDVEGLYDPNPEATG 91

Qy        537 TFYPQGGGFLHDAAHFDAGFFGISPREALAMDPQQRLLLETSWEAFERAGIDPLSVRGSR 596

Db         92 KSYAREAGFLYEAGEFDADFFGISPREALAMDPQQRLLLEASWEAFEHAGIPAATARGTS 151

Qy        597 TGVFAGALSFDYGPRMDTASSEGAADVEGHILTGTTGSVLSGRIAYSFGLEGPAITVDTG 656

Db        152 VGVFTGVMYHDYATRL-TDVPEG---IEGYLGTGNSGSVASGRVAYTLGLEGPAVTVDTA 207

Qy        657 CSASLVTLHLACQSLRSGECTLALAGGVSVMSTLGMFIEFSRQRGLSVDGRCKAYSAAAD 716

Db        208 CSSSLVALHLAVQALRKGEVDMALAGGVTVMSTPSTFVEFSRQRGLAPDGRSKSFSSTAD 267

Qy        717 GTGWGEGVGMLLVERLSDAVRLGHRVLAVVRGSAVNQDGASNGLTAPNGPAQERVIRQAL 776

Db        268 GTSWSEGVGVLLVERLSDARRKGHRILAVVRGTAVNQDGASSGLTAPNGPSQQRVIRRAL 327

Qy        777 ANAGLSVADVDVVEGHGTGTTLGDPIEAQALLATYGQ-RAGDRPLWLGSLKSNIGHTMAA 835

Db        328 ADARLTTSDVDVVEAHGTGTRLGDPIEAQAVIATYGQGRDGEQPLRLGSLKSNIGHTQAA 387

Qy        836 AGVGGVIKMVMALREGVLPRTLHVDKPSPQVDWSAGAVRLLTEAVPWPGDAAGRLRRAGV 895

Db        388 AGVSGVIKMVQAMRHGVLPKTLHVEKPTDQVDWSAGAVELLTEAMDWPDKGDGGLRRAAV 447

Qy        896 SSFGIGGTNAHVILEEAPAAGGCVAGGGVLEGAPGLAISVAESVAAPVAVSAPVAESVPV 955

Db        448 SSFGVSGTNAHVVLEEAPA----------------------------AEETPASEATPA 478

Qy        956 PVPVP----VPVPVSARSEAGLRAQAEALRQYVAVRPDVSLADVGA---GLACGRAVLEH 1008

Db        479 VEPSVGAGLVPWLVSAKTPAALDAQIGRLAAF-ASQGRTDAADPGAVARVLAGGRAEFEH 537

Qy       1009 RAVVLAADREELVQGLGA---LAAGEPDRRVTTGHAPGGDRGGVVFVFPGQGGQWAGMGV 1065

Db        538 RAVVLGTGQDDFAQALTAPEGLIRGTP----------SDVGRVAFVFPGQGTQWAGMGA 586

Qy       1066 RLLASSPVFARRMQACEEALAPWVDWSVVDILRRDAGDAVWERADVVQPVLFSVMVSLAA 1125

Db        587 ELLDVSKEFAAAMAECESALSRYVDWSLEAVVRQAPGAPTLERVDVVQPVTFAVMVSLAK 646

Qy       1126 LWRSYGIEPDAVLGHSQGEIAAAHVCGALSLKDAAKTVALRSRALAA-VRGRGGMASVPL 1184

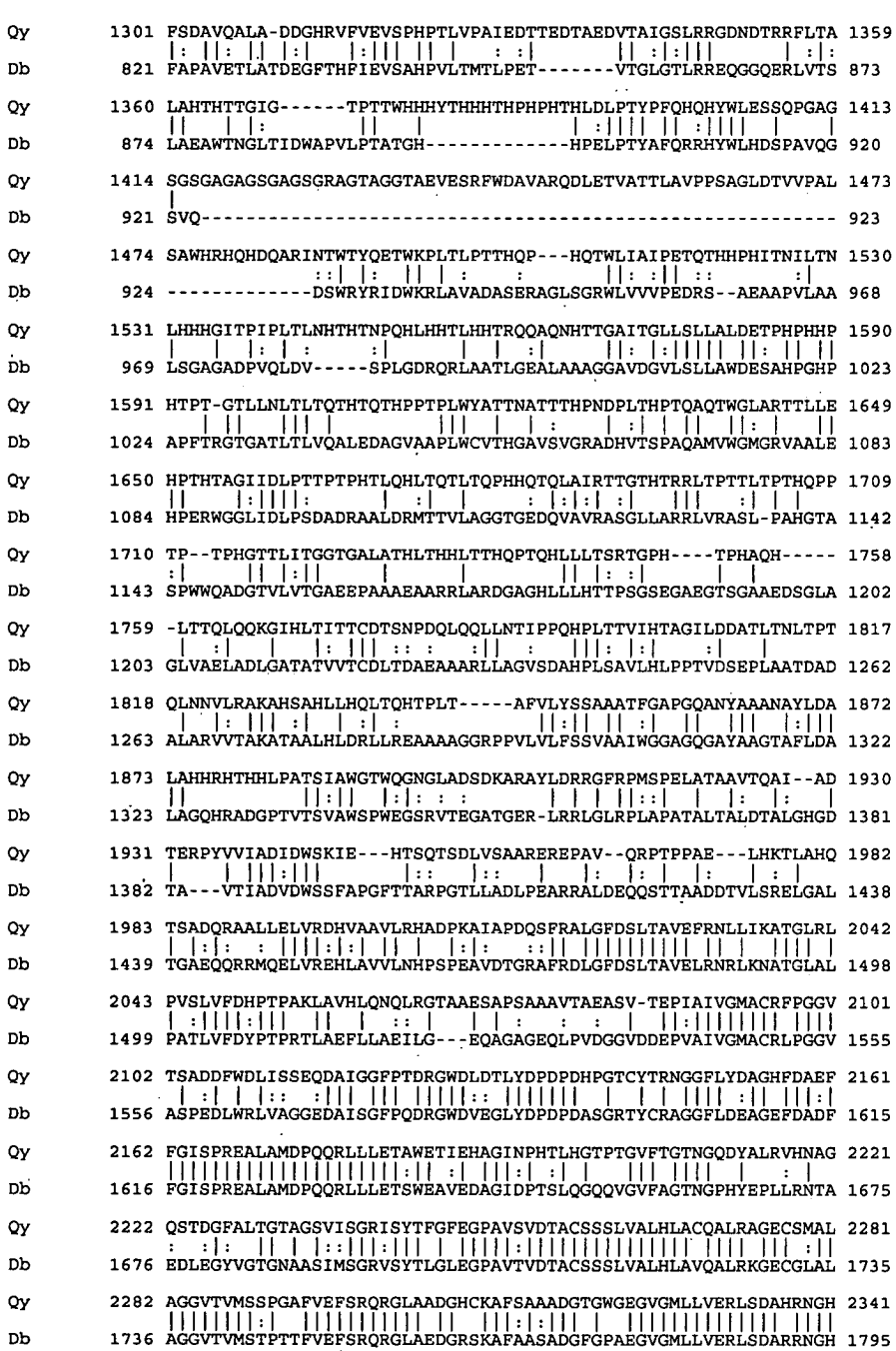
Db        647 VWQHHGVTPQAVVGHSQGEIAAAYVAGALTLDDAARVVTLRSKSIAAHLAGKGGMISLAL 706

Qy       1185 ----PAQEVEQLIGERWAGRLWVAAVNGPRSTAVSGDAEAVDEVLAYCAGTGVRARRIPV 1240

Db        707 SEEATRQRIENLHG------LSIAAVNGPTATVVSGDPTQIQELAQACEADGVRARIIPV 760

Qy       1241 DYASHCPHVQPLREELLELLGDISPQPSGVPFFSTVEGTWLDTTTLDAAYWYRNLHQPVR 1300

Db        761 DYASHSAHVETIESELAEVLAGLSPRTPEVPFFSTLEGAWITEPVLDGTYWYRNLRHRVG 820
```


Qy 1301 FSDAVQALA-DDGHRVFVEVSPHPTLVPAIEDTTEDTAEDVTAIGSLRRGDNDTRRFLTA 1359
Db 821 FAPAVETLATDEGFTHFIEVSAHPVLTMTLPET-------VTGLGTLRREQGGQERLVTS 873
Qy 1360 LAHTHTTGIG------TPTTWHHHYTHHHTHPHPHTHLDLPTYPFQHQHYWLESSQPGAG 1413
Db 874 LAEAWTNGLTIDWAPVLPTATGH-------------HPELPTYAFQRRHYWLHDSPAVQG 920
Qy 1414 SGSGAGAGSGAGSGRAGTAGGTAEVESRFWDAVARQDLETVATTLAVPPSAGLDTVVPAL 1473
Db 921 SVQ---------- 923
Qy 1474 SAWHRHQHDQARINTWTYQETWKPLTLPTTHQP---HQTWLIAIPETQTHHPHITNILTN 1530
Db 924 ------------DSWRYRIDWKRLAVADASERAGLSGRWLVVVPEDRS--AEAAPVLAA 968
Qy 1531 LHHHGITPIPLTLNHTHTNPQHLHHTLHHTRQQAQNHTTGAITGLLSLLALDETPHPHHP 1590
Db 969 LSGAGADPVQLDV-----SPLGDRQRLAATLGEALAAAGGAVDGVLSLLAWDESAHPGHP 1023
Qy 1591 HTPT-GTLLNLTLTQTHTQTHPPTPLWYATTNATTTHPNDPLTHPTQAQTWGLARTTLLE 1649
Db 1024 APFTRGTGATLTLVQALEDAGVAAPLWCVTHGAVSVGRADHVTSPAQAMVWGMGRVAALE 1083
Qy 1650 HPTHTAGIIDLPTTPTPHTLQHLTQTLTQPHHQTQLAIRTTGTHTRRLTPTTLTPTHQPP 1709
Db 1084 HPERWGGLIDLPSDADRAALDRMTTVLAGGTGEDQVAVRASGLLARRLVRASL-PAHGTA 1142
Qy 1710 TP--TPHGTTLITGGTGALATHLTHHLTTHQPTQHLLLTSRTGPH----TPHAQH----- 1758
Db 1143 SPWWQADGTVLVTGAEEPAAAEAARRLARDGAGHLLLHTTPSGSEGAEGTSGAAEDSGLA 1202
Qy 1759 -LTTQLQQKGIHLTITTCDTSNPDQLQQLLNTIPPQHPLTTVIHTAGILDDATLTNLTPT 1817
Db 1203 GLVAELADLGATATVVTCDLTDAEAAARLLAGVSDAHPLSAVLHLPPTVDSEPLAATDAD 1262
Qy 1818 QLNNVLRAKAHSAHLLHQLTQHTPLT-----AFVLYSSAAATFGAPGQANYAAANAYLDA 1872
Db 1263 ALARVVTAKATAALHLDRLLREAAAAGGRPPVLVLFSSVAAIWGGAGQGAYAAGTAFLDA 1322
Qy 1873 LAHHRHTHHLPATSIAWGTWQGNGLADSDKARAYLDRRGFRPMSPELATAAVTQAI--AD 1930
Db 1323 LAGQHRADGPTVTSVAWSPWEGSRVTEGATGER-LRRLGLRPLAPATALTALDTALGHGD 1381
Qy 1931 TERPYVVIADIDWSKIE---HTSQTSDLVSAARAREPAV--QRPTPPAE---LHKTLAHQ 1982
Db 1382 TA---VTIADVDWSSFAPGFTTARPGTLLADLPEARRALDEQQSTTAADDTVLSRELGAL 1438
Qy 1983 TSADQRAALLELVRDHVAAVLRHADPKAIAPDQSFRALGFDSLTAVEFRNLLIKATGLRL 2042
Db 1439 TGAEQQRRMQELVREHLAVVLNHPSPEAVDTGRAFRDLGFDSLTAVELRNRLKNATGLAL 1498
Qy 2043 PVSLVFDHPTPAKLAVHLQNQLRGTAAESAPSAAAVTAEASV-TEPIAIVGMACRFPGGV 2101
Db 1499 PATLVFDYPTPRTLAEFLLAEILG---EQAGAGEQLPVDGGVDDEPVAIVGMACRLPGGV 1555
Qy 2102 TSADDFWDLISSEQDAIGGFPTDRGWDLDTLYDPDPDHPGTCYTRNGGFLYDAGHFDAEF 2161
Db 1556 ASPEDLWRLVAGGEDAISGFPQDRGWDVEGLYDPDPDASGRTYCRAGGFLDEAGEFDADF 1615
Qy 2162 FGISPREALAMDPQQRLLLETAWETIEHAGINPHTLHGTPTGVFTGTNGQDYALRVHNAG 2221
Db 1616 FGISPREALAMDPQQRLLLETSWEAVEDAGIDPTSLQGQQVGVFAGTNGPHYEPLLRNTA 1675
Qy 2222 QSTDGFALTGTAGSVISGRISYTFGFEGPAVSVDTACSSSLVALHLACQALRAGECSMAL 2281
Db 1676 EDLEGYVGTGNAASIMSGRVSYTLGLEGPAVTVDTACSSSLVALHLAVQALRKGECGLAL 1735
Qy 2282 AGGVTVMSSPGAFVEFSRQRGLAADGHCKAFSAAADGTGWGEGVGMLLVERLSDAHRNGH 2341
Db 1736 AGGVTVMSTPTTFVEFSRQRGLAEDGRSKAFAASADGFGPAEGVGMLLVERLSDARRNGH 1795
Qy 2342 RVLAVVRGSAVNQDGASNGLTAPNGPSQQRVIRQALANAGLSAGDVDAVEAHGTGTTLGD 2401
Db 1796 RVLAVVRGSAVNQDGASNGLTAPNGPSQQRVIRRALADARLTTADVDVVEAHGTGTRLGD 1855
Qy 2402 PIEAQALLATYGQDRAGEGPLWLGSVKSNVGHTQAAAGVAGVIKMVMALRHGLLPRTLHV 2461
Db 1856 PIEAQALIATYGQGRDTEQPLRLGSLKSNIGHTQAAAGVSGIIKMVQAMRHGVLPKTLHV 1915
Qy 2462 DEPSPHVDWSAGAVQLLTETVPWP-GGEGRLRRAGVSSFGVSGTNAHVILEEAPAD-DVP 2519
Db 1916 DRPSDQIDWSAGTVELLTEAMDWPRKQEGGLRRAAVSSFGISGTNAHIVLEEAPVDEDAP 1975
Qy 2520 GGPPAGEGDAGSDDEAAAGSPGVWPWLVSAKSQPALRAQAQALHAHLTDHPGLDLAD--- 2576
Db 1976 A------------DEPSVG--GVVPWLVSAKTPAALDAQIGRL-AAFASQGRTDAADPGA 2020
Qy 2577 VGYTLAHARAVFDHRATLIAADRDTFLQALQALAAGEPHPAVIHSSAPGG--TGTGEAAG 2634
Db 2021 VARVLAGGRAQFEHRAVALGTGQDDLAAAL--------------AAPEGLVRGVASGVG 2065
Qy 2635 KTAFICSGQGTQRPGMAHGLYHTHPVFAAALNDICTHLDPHLDHPLLPLLTQNDNDNEDA 2694
Db 2066 RVAFVFPGQGTQWAGMGAELLDVSKEFAAAMAECEAALAPYVDWSLEAVVRQ-----APG 2120
Qy 2695 AALLQQTRYAQPALFAFQVALHRLLTDGYHITPHYYAGHSLGEITAAHLAGILTLTDATT 2754
Db 2121 APTLERVDVVQPVTFAVMVSLAK-VWQHHGVTPQAVVGHSQGEIAAAYVAGALSLDDAAR 2179
Qy 2755 LITQRATLM--QTMPPGTMTTLHTTPHHITHHLTAHENDLAIAAINTPTSLVISGTPHTV 2812
Db 2180 VVTLRSKSIGAHLAGQGGMLSLALSEAAVVERLAGFDG-LSVAAVNGPTATVVSGDPTQI 2238
Qy 2813 QHITTLCQQQGIKTKTLPTNHAFHSPHTNPILNQLHQHTQTLTYHPPHTPLITA-----N 2867
Db 2239 QELAQACEADGVRARIIPVDYASHSAHVETIESELADVLAGLSPQTPQVPFFSTLEGAWI 2298
Qy 2868 TPPDQLLTPHYWTQQARNTVDYATTTQTL-HQHGVTTYIELGPDNTLTTLTHHNLPNPPT 2926
Db 2299 TEP--ALDGGYWYRNLRHRVGFAPAVETLATDEGFTHFVEVSAHPVLTMA----LPETVT 2352
Qy 2927 TTLTLTHPHHHPQTHLLTNLAKT-----TTTWHPHHYTHHDNQPHTHTHLDLPTYPFQHH 2981
Db 2353 GLGTLRR-DNGGQHRLTTSLAEAWANGLTVDW-------ASLLPTTTTHPDLPTYAFQTE 2404
Qy 2982 HYWLESTQPGAGNVSAAGLDPTEHPLLGATLELATDGGALLAGRLSLRSHPWLADHAVGG 3041
Db 2405 RYWPQPDLSAAGDITSAGLGAAEHPLLGAAVALADSDGCLLTGSLSLRTHPWLADHAVAG 2464
Qy 3042 TVLLSGATFLELALHAGTYVGCDRVDELTLHAPLVVPVDGGVSVQVGVAAADGEGRRLVS 3101
Db 2465 TVLLPGTAFVELAFRAGDQVGCDLVEELTLDAPLVLPRRGAVRVQLSVGASDESGRRTFG 2524
Qy 3102 VYARGGSACGGGGASGGVWTCHASGVLVEAAAGGVVVDGLAGV-----WPPRGAVAVDVD 3156
Db 2525 LYAHPEDAPG-----EAEWTRHATGVLAARA------DRTAPVADPEAWPPPGAEPVDVD 2573
Qy 3157 GVRDRLAGAGCVLGPVFSGLRAVWRDGGDLLAEVCLPEEAWG-DAAGFGLHPALLDGVVQ 3215
Db 2574 GLYERFAANGYGYGPLFQGVRGVWRRGDEVFADVALPAEVAGAEGARFGLHPALLDAAVQ 2633
Qy 3216 PLSVLLPGGTGFGEGAGFGEGVRVPAVWGGVSLHRAGVTGVRVRVSAVGRGGGREAVSVV 3275
Db 2634 ------AAGAGRGVRRGHAAAVRLER----DLLYAVGATALRVRLAP----AGPDTVSVS 2679
Qy 3276 VGDEAGVPVASVDRLELRPVDMGQLRAVSVSAGRRGSLYAVQW--------AEVGPVPVC 3327
Db 2680 AADSSGQPVFAADSLTVLPVDPAQLAAFSDPT--LDALHLLEWTAWDGAAQALPGAVVLG 2737
Qy 3328 GQA----------------WAWHEDVGES---GGGPVPGVVVLRCPDAGAGGGGGGGGGG 3368
Db 2738 GDADGLAAALRAGGTEVLSFPDLTDLVEAVDRGETPAPATVLVACPAAGP------DGPE 2791
Qy 3369 GVGEVVGGVLGVVQGWLGLERFAGSRLVVVTRGAVVAGPEDGPVDVVGASVWGLVRSAQA 3428

```
Db       2792 HVREALHGSLALMQAWLADERFTDGRLVLVTRDAVAARSGDGLRSTGQAAVWGLGRSAQT 2851

Qy       3429 EHPDRFVLLDL--------DTDTGTDLDTGAGAGWGVD-----GGRVAAVVACGEPQLAV 3475

Db       2852 ESPGRFVLLDLAGEARTAGDATAGDGLTTGDATVGGTSGDAALGSALATALGSGEPQLAL 2911

Qy       3476 RGERLLAARLKR-----------------------------------LES----SGD- 3493

Db       2912 RDGALLVPRLARAAAPAAADGLAAADGLAALPLPAAPALWRLEPGTDGSLESLTAAPGDA 2971

Qy       3494 ------------------------------------------------------VP 3495

Db       2972 ETLAPEPLGPGQVRIAIRATGLNFRDVLIALGMYPDPALMGTEGAGVVTATGPGVTHLAP 3031

Qy       3496 AQR-----------------------SGDTRARRSDVPA-------------------- 3511

Db       3032 GDRVMGLLSGAYAPVVVADARTVARMPEGWTFAQGASVPVVFLTAVYALRDLADVKPGER 3091

Qy       3512 ----QRSGGVP------------------------------------------------- 3518

Db       3092 LLVHSAAGGVGMAAVQLARHWGVEVHGTASHGKWDALRALGLDDAHIASSRTLDFESAFR 3151

Qy       3519 ---------------ARRSVDVS----------------------------------- 3526

Db       3152 AASGGAGMDVVLNSLAREFVDASLRLLGPGGRFVEMGKTDVRDAERVAADHPGVGYRAFD 3211

Qy       3527 ------------------------------------------------GREVLPWLS 3535

Db       3212 LGEAGPERIGEMLAEVIALFEDGVLRHLPVTTWDVRRARDAFRHVSQARHTGKVVLTMPS 3271

Qy       3536 G----GSVLVTGGTGVLGAAVARHLAGVCGVRDLLLVSRRGPDAPGAEGLRAELAALGAE 3591

Db       3272 GLDPEGTVLLTGGTGALGGIVARHVVGEWGVRRLLLVSRRGTDAPGAGELVHELEALGAD 3331

Qy       3592 VRIVACDVGERREVVRLLEGVPAGCPLTGVVHAAGVLDDATIASLTPERLGTVFAAKVDA 3651

Db       3332 VSVAACDVADREALTAVLDSIPAEHPLTAVVHTAGVLSDGTLPSMTAEDVEHVLRPKVDA 3391

Qy       3652 ALLLDELTR--GMELSAFVLFSSAAGILGSAGQGNYAAANAALDALAYRRRAAGLPGVSL 3709

Db       3392 AFLLDELTSTPGYDLAAFVMFSSAAAVFGGAGQGAYAAANATLDALAWRRRTAGLPALSL 3451

Qy       3710 AWGLWEEASGMTGHLAGTDHRRIIRSGLHPMSTPDALALFDAALALDRPVLLPADLRPAP 3769

Db       3452 GWGLWAETSGMTGGLSDTDRSRLARSGATPMDSELTLSLLDAAMRRDDPALVPIALDVAA 3511

Qy       3770 --------PLPPLLQDL--------LPATRRRTTRTTTTGGA-DNGAQLHARLAGQTHEQ 3812

Db       3512 LRAQQRDGMLAPLLSGLTRGSRVGGAPVNQRR----AAAGGAGEADTDLGGRLAAMTPDD 3567

Qy       3813 QHTTLLALVRSHIATVLGHTTPDTIPPDRAFRDLGFDSLTAVELRNRLSRTTGLRLPTTL 3872

Db       3568 RVAHLRDLVRTHVATVLGHGTPSRVDLERAFRDTGFDSLTAVELRNRLNAATGLRLPATL 3627

Qy       3873 AFDHPNPTTLTHHLHTQL-----------------QPQPDNAVAPVLAELDKLESALSAL 3915

Db       3628 VFDHPTPGELAGHLLDELATAAGGSWAEGTGSGDTASATDRQTTAALAELDRLEGVLASL 3687

Qy       3916 DKTDSASERVTLRLKSLMLRWNAPQHPTAESADDDEKFTSATEAEIFKFIDNDLG 3970

Db       3688 APAAGGRPELAARLRALAA-------ALGDDGDDATDLDEASDDDLFSFIDKELG 3735


RESULT 15
US-09-861-289-33
; Sequence 33, Application US/09861289
; Patent No. US20020110897A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/861,289
;  CURRENT FILING DATE:  2001-05-18
;  PRIOR APPLICATION NUMBER: 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 33
;   LENGTH: 3739
;   TYPE: PRT
;   ORGANISM: Streptomyces venezuelae
US-09-861-289-33

  Query Match             33.7%;  Score 7017;  DB 10;  Length 3739;
  Best Local Similarity   42.4%;  Pred. No. 0;
  Matches 1676;  Conservative  390;  Mismatches 1177;  Indels  712;  Gaps   82;

Qy        477 AAADDPIAIIGMACRFPGGVRSPKDLWELAASGGDAIGPFPTDRGWPTEQRHAQDPTQPG 536

Db         32 AKAGEPVAIVGMACRLPGGVASPEDLWRLVAGGEDAISEFPQDRGWDVEGLYDPNPEATG 91

Qy        537 TFYPQGGGFLHDAAHFDAGFFGISPREALAMDPQQRLLLETSWEAFERAGIDPLSVRGSR 596

Db         92 KSYAREAGFLYEAGEFDADFFGISPREALAMDPQQRLLLEASWEAFEHAGIPAATARGTS 151

Qy        597 TGVFAGALSFDYGPRMDTASSEGAADVEGHILTGTTGSVLSGRIAYSFGLEGPAITVDTG 656

Db        152 VGVFTGVMYHDYATRL-TDVPEG---IEGYLGTGNSGSVASGRVAYTLGLEGPAVTVDTA 207

Qy        657 CSASLVTLHLACQSLRSGECTLALAGGVSVMSTLGMFIEFSRQRGLSVDGRCKAYSAAAD 716

Db        208 CSSSLVALHLAVQALRKGEVDMALAGGVTVMSTPSTFVEFSRQRGLAPDGRSKSFSSTAD 267

Qy        717 GTGWGEGVGMLLVERLSDAVRLGHRVLAVVRGSAVNQDGASNGLTAPNGPAQERVIRQAL 776

Db        268 GTSWSEGVGVLLVERLSDARRKGHRILAVVRGTAVNQDGASSGLTAPNGPSQQRVIRRAL 327

Qy        777 ANAGLSVADVDVVEGHGTGTTLGDPIEAQALLATYGQ-RAGDRPLWLGSLKSNIGHTMAA 835

Db        328 ADARLTTSDVDVVEAHGTGTRLGDPIEAQAVIATYGQGRDGEQPLRLGSLKSNIGHTQAA 387

Qy        836 AGVGGVIKMVMALREGVLPRTLHVDKPSPQVDWSAGAVRLLTEAVPWPGDAAGRLRRAGV 895

Db        388 AGVSGVIKMVQAMRHGVLPKTLHVEKPTDQVDWSAGAVELLTEAMDWPDKGDGGLRRAAV 447

Qy        896 SSFGIGGTNAHVILEEAPAAGGCVAGGGVLEGAPGLAISVAESVAAPVAVSAPVAESVPV 955

Db        448 SSFGVSGTNAHVVLEEAPA----------------------------AEETPASEATPA 478

Qy        956 PVPVP----VPVPVSARSEAGLRAQAEALRQYVAVRPDVSLADVGA---GLACGRAVLEH 1008

Db        479 VEPSVGAGLVPWLVSAKTPAALDAQIGRLAAF-ASQGRTDAADPGAVARVLAGGRAEFEH 537

Qy       1009 RAVVLAADREELVQGLGA---LAAGEPDRRVTTGHAPGGDRGGVVFVFPGQGGQWAGMGV 1065

Db        538 RAVVLGTGQDDFAQALTAPEGLIRGTP-----------SDVGRVAFVFPGQGTQWAGMGA 586

Qy       1066 RLLASSPVFARRMQACEEALAPWVDWSVVDILRRDAGDAVWERADVVQPVLFSVMVSLAA 1125

Db        587 ELLDVSKEFAAAMAECESALSRYVDWSLEAVVRQAPGAPTLERVDVVQPVTFAVMVSLAK 646

Qy       1126 LWRSYGIEPDAVLGHSQGEIAAAHVCGALSLKDAAKTVALRSRALAA-VRGRGGMASVPL 1184

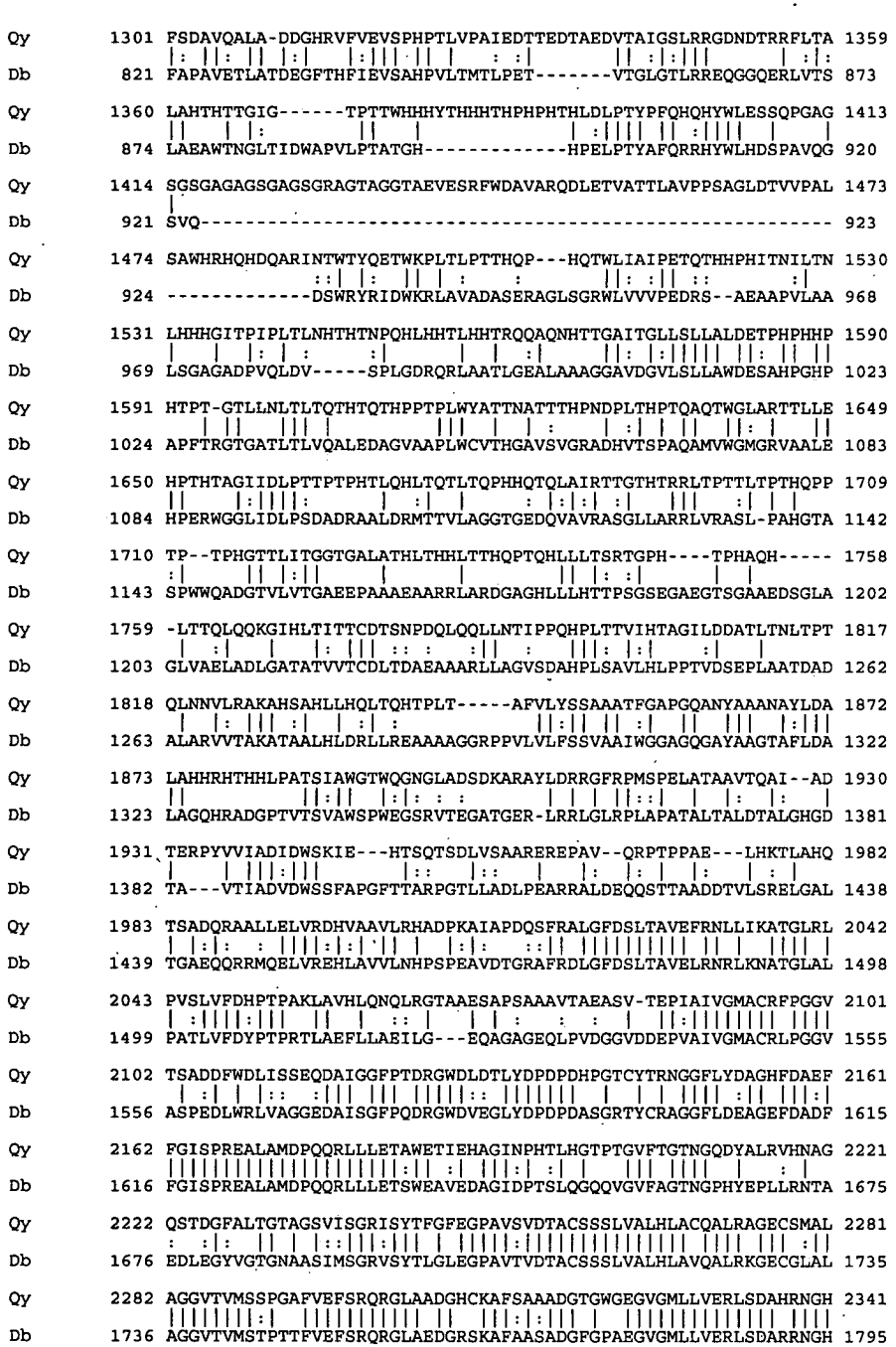
Db        647 VWQHHGVTPQAVVGHSQGEIAAAYVAGALTLDDAARVVTLRSKSIAAHLAGKGGMISLAL 706

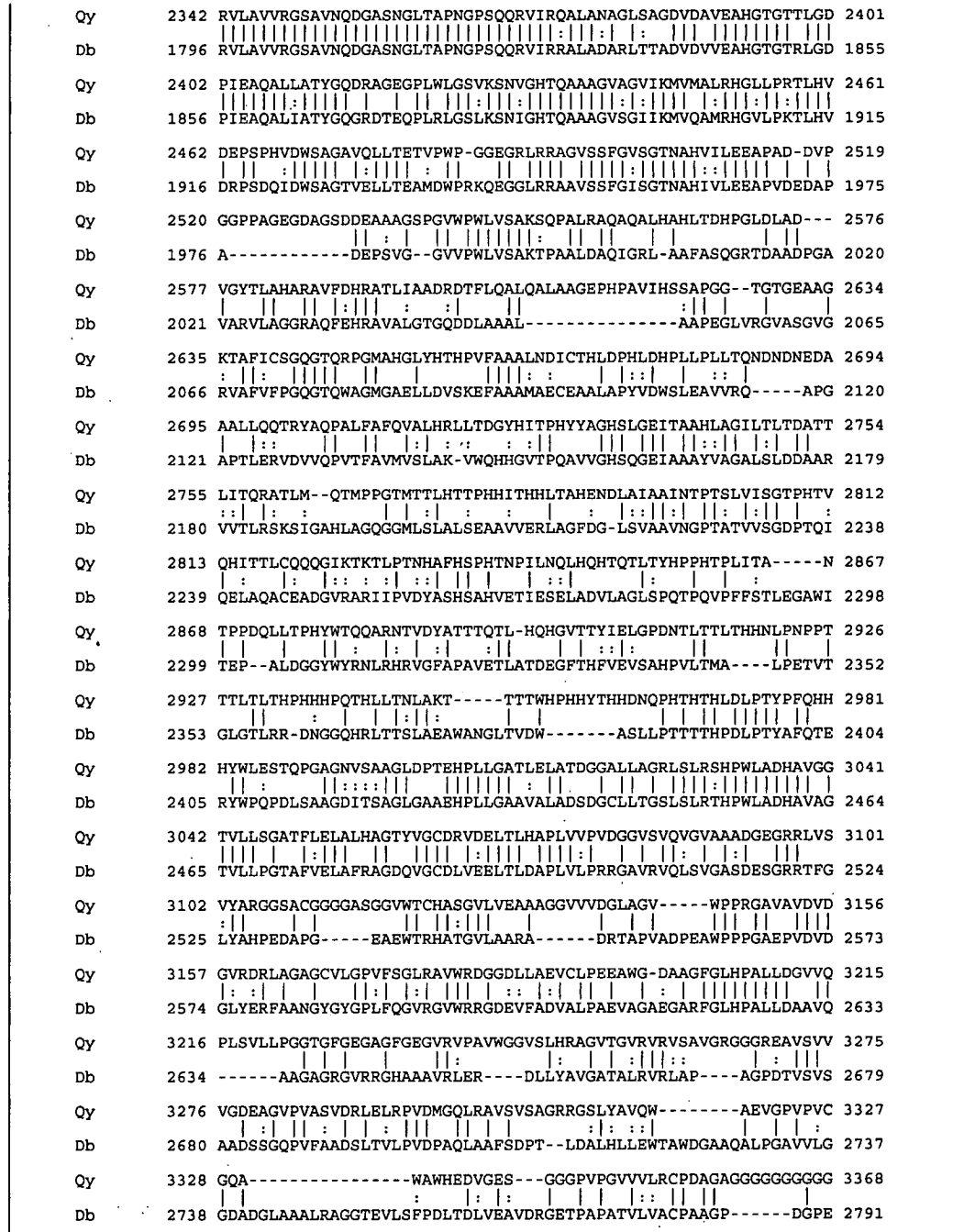
Qy       1185 ----PAQEVEQLIGERWAGRLWVAAVNGPRSTAVSGDAEAVDEVLAYCAGTGVRARRIPV 1240

Db        707 SEEATRQRIENLHG------LSIAAVNGPTATVVSGDPTQIQELAQACEADGVRARIIPV 760

Qy       1241 DYASHCPHVQPLREELLELLGDISPQPSGVPFFSTVEGTWLDTTTLDAAYWYRNLHQPVR 1300

Db        761 DYASHSAHVETIESELAEVLAGLSPRTPEVPFFSTLEGAWITEPVLDGTYWYRNLRHRVG 820
```





```
Db       2792 HVREALHGSLALMQAWLADERFTDGRLVLVTRDAVAARSGDGLRSTGQAAVWGLGRSAQT 2851

Qy       3429 EHPDRFVLLDL--------DTDTGTDLDTGAGAGWGVD-----GGRVAAVVACGEPQLAV 3475

Db       2852 ESPGRFVLLDLAGEARTAGDATAGDGLTTGDATVGGTSGDAALGSALATALGSGEPQLAL 2911

Qy       3476 RGERLLAARLKR-------------------------------------LES----SGD- 3493

Db       2912 RDGALLVPRLARAAAPAAADGLAAADGLAALPLPAAPALWRLEPGTDGSLESLTAAPGDA 2971

Qy       3494 ------------------------------------------------------VP 3495

Db       2972 ETLAPEPLGPGQVRIAIRATGLNFRDVLIALGMYPDPALMGTEGAGVVTATGPGVTHLAP 3031

Qy       3496 AQR------------------------SGDTRARRSDVPA-------------------- 3511

Db       3032 GDRVMGLLSGAYAPVVVADARTVARMPEGWTFAQGASVPVVFLTAVYALRDLADVKPGER 3091

Qy       3512 ----QRSGGVP---------------------------------------------------- 3518

Db       3092 LLVHSAAGGVGMAAVQLARHWGVEVHGTASHGKWDALRALGLDDAHIASSRTLDFESAFR 3151

Qy       3519 ---------------ARRSVDVS------------------------------------ 3526

Db       3152 AASGGAGMDVVLNSLAREFVDASLRLLGPGGRFVEMGKTDVRDAERVAADHPGVGYRAFD 3211

Qy       3527 ------------------------------------------------GREVLPWLS 3535

Db       3212 LGEAGPERIGEMLAEVIALFEDGVLRHLPVTTWDVRRARDAFRHVSQARHTGKVVLTMPS 3271

Qy       3536 G----GSVLVTGGTGVLGAAVARHLAGVCGVRDLLLVSRRGPDAPGAEGLRAELAALGAE 3591

Db       3272 GLDPEGTVLLTGGTGALGGIVARHVVGEWGVRRLLLVSRRGTDAPGAGELVHELEALGAD 3331

Qy       3592 VRIVACDVGERREVVRLLEGVPAGCPLTGVVHAAGVLDDATIASLTPERLGTVFAAKVDA 3651

Db       3332 VSVAACDVADREALTAVLDSIPAEHPLTAVVHTAGVLSDGTLPSMTAEDVEHVLRPKVDA 3391

Qy       3652 ALLLDELTR--GMELSAFVLFSSAAGILGSAGQGNYAAANAALDALAYRRRAAGLPGVSL 3709

Db       3392 AFLLDELTSTPGYDLAAFVMFSSAAAVFGGAGQGAYAAANATLDALAWRRRTAGLPALSL 3451

Qy       3710 AWGLWEEASGMTGHLAGTDHRRIIRSGLHPMSTPDALALFDAALALDRPVLLPADLRPAP 3769

Db       3452 GWGLWAETSGMTGGLSDTDRSRLARSGATPMDSELTLSLLDAAMRRDDPALVPIALDVAA 3511

Qy       3770 --------PLPPLLQDL--------LPATRRRTTRTTTTGGA-DNGAQLHARLAGQTHEQ 3812

Db       3512 LRAQQRDGMLAPLLSGLTRGSRVGGAPVNQRR----AAAGGAGEADTDLGGRLAAMTPDD 3567

Qy       3813 QHTTLLALVRSHIATVLGHTTPDTIPPDRAFRDLGFDSLTAVELRNRLSRTTGLRLPTTL 3872

Db       3568 RVAHLRDLVRTHVATVLGHGTPSRVDLERAFRDTGFDSLTAVELRNRLNAATGLRLPATL 3627

Qy       3873 AFDHPNPTTLTHHLHTQL-----------------QPQPDNAVAPVLAELDKLESALSAL 3915

Db       3628 VFDHPTPGELAGHLLDELATAAGGSWAEGTGSGDTASATDRQTTAALAELDRLEGVLASL 3687

Qy       3916 DKTDSASERVTLRLKSLMLRWNAPQHPTAESADDDEKFTSATEAEIFKFIDNDLG 3970

Db       3688 APAAGGRPELAARLRALAA-------ALGDDGDDATDLDEASDDDLFSFIDKELG 3735
```

Search completed: June 17, 2003, 13:46:38
Job time : 197.76 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 17, 2003, 12:55:32 ; Search time 88.6945 Seconds
(without alignments)
4305.183 Million cell updates/sec

Title: US-09-914-286-3
Perfect score: 20829
Sequence: 1 VQRMDGGEEPRPAAGEVLGV.....FTSATEAEIFKFIDNDLGLS 3972

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_73:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	8148	39.1	4613	2	T17409	polyketide synthas
2	7098	34.1	5069	2	T17464	rifamycin polyketi
3	7017	33.7	3739	2	T17410	polyketide synthas
4	6813.5	32.7	4151	2	G70944	probable polyketid
5	6756	32.4	10223	2	T30225	polyketide synthas
6	6642.5	31.9	8563	2	T30226	polyketide synthas
7	6416	30.8	7576	2	T17428	FK506 polyketide s
8	6360	30.5	3491	2	T43231	probable 6-deoxyer
9	6293.5	30.2	3573	2	S23070	erythronolide synt
10	6164.5	29.6	6260	2	T30228	polyketide synthas
11	6108.5	29.3	3519	2	S43048	polyketide synthas
12	6077.5	29.2	3413	2	T17467	rifamycin polyketi
13	6068.5	29.1	4735	2	T17463	rifamycin polyketi
14	5922	28.4	6420	2	T30283	polyketide synthas
15	5392	25.9	3172	2	S22012	erythronolide synt
16	5306	25.5	3178	2	S13595	6-deoxyerythronoli
17	3744.5	18.0	2723	2	T03221	probable polyketid
18	3710	17.8	1762	2	T03222	probable polyketid
19	3555	17.1	1763	2	T17465	rifamycin polyketi
20	3549	17.0	2103	2	G86925	probable polyketid
21	3470.5	16.7	1937	2	T03224	probable polyketid
22	3461.5	16.6	1728	2	T17466	rifamycin polyketi
23	3416	16.4	2126	2	H70621	probable polyketid
24	3415	16.4	2100	2	T03223	probable polyketid
25	3372	16.2	2124	2	T28658	polyketide synthas
26	3280.5	15.7	1562	2	T17411	polyketide synthas
27	2967.5	14.2	1602	2	H70984	probable polyketid
28	2461	11.8	2458	2	T17420	probable polyketid
29	2428	11.7	8243	2	T31307	type I fatty acid
30	2412.5	11.6	1346	2	T17412	polyketide synthas
31	2256.5	10.8	4930	2	E69679	polyketide synthet
32	2142.5	10.3	1827	2	B70984	probable polyketid
33	2097.5	10.1	2478	2	AH2140	polyketide synthas
34	2079	10.0	1570	2	AC2012	hypothetical prote
35	2060	9.9	2518	2	AI2140	polyketide synthas
36	2050.5	9.8	1538	2	E70874	probable ppsB prot
37	2042	9.8	4427	2	PN0637	polyketide synthas
38	2040	9.8	2188	2	A70984	probable polyketid
39	2030	9.7	1616	2	G70668	polyketide synthas
40	2017	9.7	1876	2	C70749	probable ppsA prot
41	2015.5	9.7	1540	2	H87203	polyketide synthas
42	2015	9.7	1815	2	S73021	polyketide synthas
43	2015	9.7	1822	2	F87203	polyketide synthas
44	1969.5	9.5	2201	2	S73014	polyketide synthas
45	1962.5	9.4	1871	2	A87204	polyketide synthas

ALIGNMENTS

RESULT 1
T17409
polyketide synthase type I - Streptomyces venezuelae
C;Species: Streptomyces venezuelae
C;Date: 02-Sep-2000 #sequence_revision 02-Sep-2000 #text_change 17-Nov-2000
C;Accession: T17409
R;Xue, Y.; Zhao, L.; Liu, H.W.; Sherman, D.H.
Proc. Natl. Acad. Sci. U.S.A. 95, 12111-12116, 1998
A;Title: A gene cluster for macrolide antibiotic biosynthesis in streptomyces venezuela
A;Reference number: Z18773; MUID:98445333; PMID:9770448
A;Accession: T17409
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-4613 <XUE>
A;Cross-references: EMBL:AF079138; NID:g3808326; PID:g3800834; PIDN:AAC69329.1
C;Genetics:
A;Gene: pikAI
C;Superfamily: acyl carrier protein homology
C;Keywords: antibiotic biosynthesis; carrier protein
F;1010-1081/Domain: acyl carrier protein homology <ACP1>
F;2495-2566/Domain: acyl carrier protein homology <ACP2>
F;4407-4478/Domain: acyl carrier protein homology <ACP3>

Query Match 39.1%; Score 8148; DB 2; Length 4613;
Best Local Similarity 44.8%; Pred. No. 0;
Matches 1909; Conservative 463; Mismatches 1307; Indels 586; Gaps 94;

```
Qy        15 GEVLGVADEADGGVVFVFPGQGPQWPGMGRELLDASDVFRESVRACEAAFAPYVDWSVEQ 74
             | | |||     | | ||||||| || ||| ||||:| ||  ::  |||| :||||||:|
Db       620 GLVRGVASGV-GRVAFVFPGQGTQWAGMGAELLDSSAVFAAAMAECEAALSPYVDWSLEA 678

Qy        75 VLRDSPDAPGLDRVDVVQPTLFAVMISLAALWRSQGVEPCAVLGHSLGEIAAAHVSGGLS 134
             |:| :| || |:|||||||  ||||:||| :|:  || | ||:||| ||||||:|:| ||
Db       679 VVRQAPGAPTLERVDVVQPVTFAVMVSLARVWQHHGVTPQAVVGHSQGEIAAAYVAGALS 738

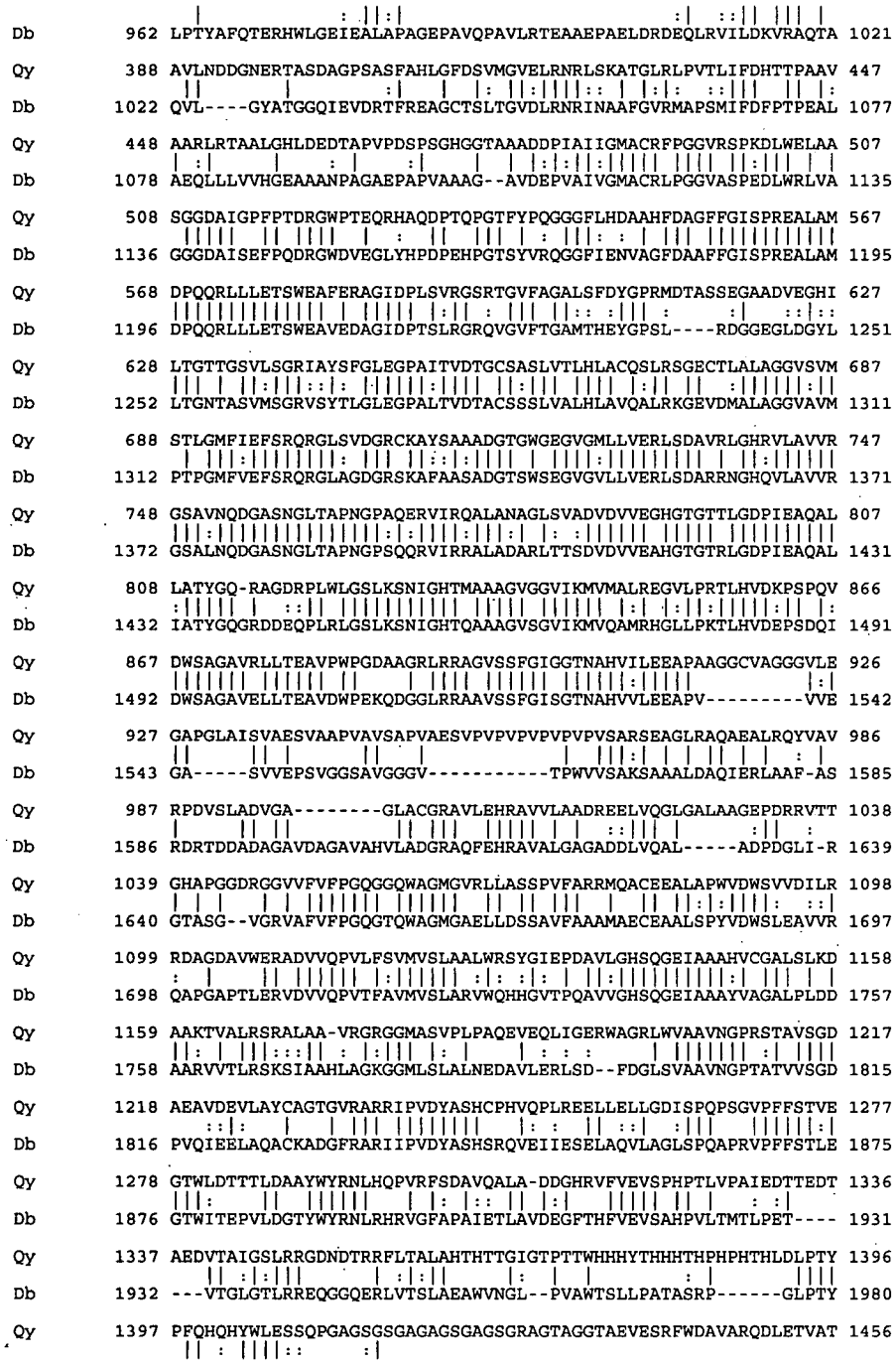
Qy       135 LADAARVVTLWSQA-QTTLAGTGALVSVAATPDELLPRIAPWTEDNPARLAVAAVNGPRS 193
             | |||||||| |::    ||| | ::|:| : | :| |:| :       |:|||||| :
Db       739 LDDAARVVTLRSKSIAAHLAGKGGMLSLALSEDAVLERLAGFD-----GLSVAAVNGPTA 793

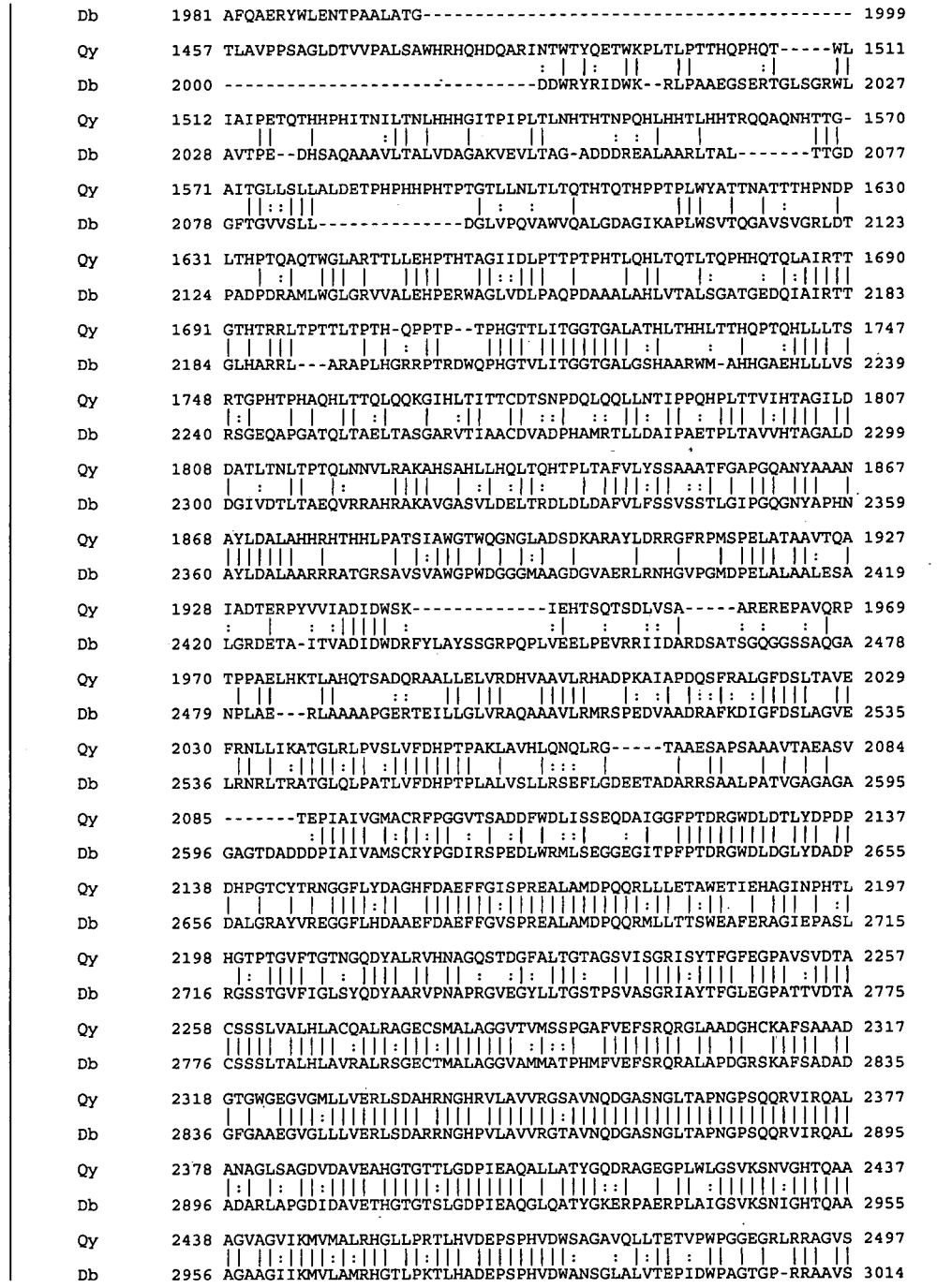
Qy       194 TVVSGAREAVADLVADLTAAQVRTRMIPVDVPAHSPLMYAIEERVVSGLLPITPRPSRIP 253
             |||||    : :|      |  || |:|||  :||  :  ||  :   |  ::|:  |:|
Db       794 TVVSGDPVQIEELARACEADGVRARVIPVDYASHSRQVEIIESELAEVLAGLSPQAPRVP 853

Qy       254 FHSSVTGGRLDTRELDAAYWYRNMSSTVRFEPAAR-LLLQQGPKTFVEMSPHPVLTMGLQ 312
             | |:: |  :    ||  |||||:   | | ||   |   :|   |||:| |||||| |
Db       854 FFSTLEGAWITEPVLDGGYWYRNLRHRVGFAPAVETLATDEGFTHFVEVSAHPVLTMALP 913

Qy       313 ELAPDLGDTTGTADTVIMGTLRRGQGTLDHFLTSLAQ---------------LRGHGET 356
                 |  || |      ||||  |  | : |||:                  ||
Db       914 ------GTVTGLA------TLRRDNGGQDRLVASLAEAWANGLAVDWSPLLPSATGHHSD 961

Qy       357 SATTVLS------ARLTALSPT----------------------QQQSLLLDLVRAHTM 387
```


Db 962 LPTYAFQTERHWLGEIEALAPAGEPAVQPAVLRTEAAEPAELDRDEQLRVILDKVRAQTA 1021
Qy 388 AVLNDDGNERTASDAGPSASFAHLGFDSVMGVELRNRLSKATGLRLPVTLIFDHTTPAAV 447
Db 1022 QVL----GYATGGQIEVDRTFREAGCTSLTGVDLRNRINAAFGVRMAPSMIFDFPTPEAL 1077
Qy 448 AARLRTAALGHLDEDTAPVPDSPSGHGGTAAADDPIAIIGMACRFPGGVRSPKDLWELAA 507
Db 1078 AEQLLLVVHGEAAANPAGAEPAPVAAAG--AVDEPVAIVGMACRLPGGVASPEDLWRLVA 1135
Qy 508 SGGDAIGPFPTDRGWPTEQRHAQDPTQPGTFYPQGGGFLHDAAHFDAGFFGISPREALAM 567
Db 1136 GGGDAISEFPQDRGWDVEGLYHPDPEHPGTSYVRQGGFIENVAGFDAAFFGISPREALAM 1195
Qy 568 DPQQRLLLETSWEAFERAGIDPLSVRGSRTGVFAGALSFDYGPRMDTASSEGAADVEGHI 627
Db 1196 DPQQRLLLETSWEAVEDAGIDPTSLRGRQVGVFTGAMTHEYGPSL----RDGGEGLDGYL 1251
Qy 628 LTGTTGSVLSGRIAYSFGLEGPAITVDTGCSASLVTLHLACQSLRSGECTLALAGGVSVM 687
Db 1252 LTGNTASVMSGRVSYTLGLEGPALTVDTACSSSLVALHLAVQALRKGEVDMALAGGVAVM 1311
Qy 688 STLGMFIEFSRQRGLSVDGRCKAYSAAADGTGWGEGVGMLLVERLSDAVRLGHRVLAVVR 747
Db 1312 PTPGMFVEFSRQRGLAGDGRSKAFAASADGTSWSEGVGVLLVERLSDARRNGHQVLAVVR 1371
Qy 748 GSAVNQDGASNGLTAPNGPAQERVIRQALANAGLSVADVDVVEGHGTGTTLGDPIEAQAL 807
Db 1372 GSALNQDGASNGLTAPNGPSQQRVIRRALADARLTTSDVDVVEAHGTGTRLGDPIEAQAL 1431
Qy 808 LATYGQ-RAGDRPLWLGSLKSNIGHTMAAAGVGGVIKMVMALREGVLPRTLHVDKPSPQV 866
Db 1432 IATYGQGRDDEQPLRLGSLKSNIGHTQAAAGVSGVIKMVQAMRHGLLPKTLHVDEPSDQI 1491
Qy 867 DWSAGAVRLLTEAVPWPGDAAGRLRRAGVSSFGIGGTNAHVILEEAPAAGGCVAGGGVLE 926
Db 1492 DWSAGAVELLTEAVDWPEKQDGGLRRAAVSSFGISGTNAHVVLEEAPV---------VVE 1542
Qy 927 GAPGLAISVAESVAAPVAVSAPVAESVPVPVPVPVPVPVSARSEAGLRAQAEALRQYVAV 986
Db 1543 GA-----SVVEPSVGGSAVGGGV-----------TPWVVSAKSAAALDAQIERLAAF-AS 1585
Qy 987 RPDVSLADVGA--------GLACGRAVLEHRAVVLAADREELVQGLGALAAGEPDRRVTT 1038
Db 1586 RDRTDDADAGAVDAGAVAHVLADGRAQFEHRAVALGAGADDLVQAL-----ADPDGLI-R 1639
Qy 1039 GHAPGGDRGGVVFVFPGQGGQWAGMGVRLLASSPVFARRMQACEEALAPWVDWSVVDILR 1098
Db 1640 GTASG--VGRVAFVFPGQGTQWAGMGAELLDSSAVFAAAMAECEAALSPYVDWSLEAVVR 1697
Qy 1099 RDAGDAVWERADVVQPVLFSVMVSLAALWRSYGIEPDAVLGHSQGEIAAAHVCGALSLKD 1158
Db 1698 QAPGAPTLERVDVVQPVTFAVMVSLARVWQHHGVTPQAVVGHSQGEIAAAYVAGALPLDD 1757
Qy 1159 AAKTVALRSRALAA-VRGRGGMASVPLPAQEVEQLIGERWAGRLWVAAVNGPRSTAVSGD 1217
Db 1758 AARVVTLRSKSIAAHLAGKGGMLSLALNEDAVLERLSD--FDGLSVAAVNGPTATVVSGD 1815
Qy 1218 AEAVDEVLAYCAGTGVRARRIPVDYASHCPHVQPLREELLELLGDISPQPSGVPFFSTVE 1277
Db 1816 PVQIEELAQACKADGFRARIIPVDYASHSRQVEIIESELAQVLAGLSPQAPRVPFFSTLE 1875
Qy 1278 GTWLDTTTLDAAYWYRNLHQPVRFSDAVQALA-DDGHRVFVEVSPHPTLVPAIEDTTEDT 1336
Db 1876 GTWITEPVLDGTYWYRNLRHRVGFAPAIETLAVDEGFTHFVEVSAHPVLTMTLPET---- 1931
Qy 1337 AEDVTAIGSLRRGDNDTRRFLTALAHTHTTGIGTPTTWHHHYTHHHTHPHPHTHLDLPTY 1396
Db 1932 ---VTGLGTLRREQGGQERLVTSLAEAWVNGL--PVAWTSLLPATASRP------GLPTY 1980
Qy 1397 PFQHQHYWLESSQPGAGSGSGAGAGSGAGSGRAGTAGGTAEVESRFWDAVARQDLETVAT 1456


Db 1981 AFQAERYWLENTPAALATG---------- 1999
Qy 1457 TLAVPPSAGLDTVVPALSAWHRHQHDQARINTWTYQETWKPLTLPTTHQPHQT-----WL 1511
Db 2000 ----------DDWRYRIDWK--RLPAAEGSERTGLSGRWL 2027
Qy 1512 IAIPETQTHHPHITNILTNLHHHGITPIPLTLNHTHTNPQHLHHTLHHTRQQAQNHTTG- 1570
Db 2028 AVTPE--DHSAQAAAVLTALVDAGAKVEVLTAG-ADDDREALAARLTAL-------TTGD 2077
Qy 1571 AITGLLSLLALDETPHPHHPHTPTGTLLNLTLTQTHTQTHPPTPLWYATTNATTTHPNDP 1630
Db 2078 GFTGVVSLL--------------DGLVPQVAWVQALGDAGIKAPLWSVTQGAVSVGRLDT 2123
Qy 1631 LTHPTQAQTWGLARTTLLEHPTHTAGIIDLPTTPTPHTLQHLTQTLTQPHHQTQLAIRTT 1690
Db 2124 PADPDRAMLWGLGRVVALEHPERWAGLVDLPAQPDAAALAHLVTALSGATGEDQIAIRTT 2183
Qy 1691 GTHTRRLTPTTLTPTH-QPPTP--TPHGTTLITGGTGALATHLTHHLTTHQPTQHLLLTS 1747
Db 2184 GLHARRL---ARAPLHGRRPTRDWQPHGTVLITGGTGALGSHAARWM-AHHGAEHLLLVS 2239
Qy 1748 RTGPHTPHAQHLTTQLQQKGIHLTITTCDTSNPDQLQQLLNTIPPQHPLTTVIHTAGILD 1807
Db 2240 RSGEQAPGATQLTAELTASGARVTIAACDVADPHAMRTLLDAIPAETPLTAVVHTAGALD 2299
Qy 1808 DATLTNLTPTQLNNVLRAKAHSAHLLHQLTQHTPLTAFVLYSSAAATFGAPGQANYAAAN 1867
Db 2300 DGIVDTLTAEQVRRAHRAKAVGASVLDELTRDLDLDAFVLFSSVSSTLGIPGQGNYAPHN 2359
Qy 1868 AYLDALAHHRHTHHLPATSIAWGTWQGNGLADSDKARAYLDRRGFRPMSPELATAAVTQA 1927
Db 2360 AYLDALAAARRRATGRSAVSVAWGPWDGGGMAAGDGVAERLRNHGVPGMDPELALAALESA 2419
Qy 1928 IADTERPYVVIADIDWSK-------------IEHTSQTSDLVSA-----ARE REPAVQRP 1969
Db 2420 LGRDETA-ITVADIDWDRFYLAYSSGRPQPLVEELPEVRRIIDARDSATSGQGGSSAQGA 2478
Qy 1970 TPPAELHKTLAHQTSADQRAALLELVRDHVAAVLRHADPKAIAPDQSFRALGFDSLTAVE 2029
Db 2479 NPLAE---RLAAAAPGERTEILLGLVRAQAAAVLRMRSPEDVAADRAFKDIGFDSLAGVE 2535
Qy 2030 FRNLLIKATGLRLPVSLVFDHPTPAKLAVHLQNQLRG-----TAAESAPSAAAVTAEASV 2084
Db 2536 LRNRLTRATGLQLPATLVFDHPTPLALVSLLRSEFLGDEETADARRSAALPATVGAGAGA 2595
Qy 2085 -------TEPIAIVGMACRFPGGVTSADDFWDLISSEQDAIGGFPTDRGWDLDTLYDPDP 2137
Db 2596 GAGTDADDDPIAIVAMSCRYPGDIRSPEDLWRMLSEGGEGITPFPTDRGWDLDGLYDADP 2655
Qy 2138 DHPGTCYTRNGGFLYDAGHFDAEFFGISPREALAMDPQQRLLLETAWETIEHAGINPHTL 2197
Db 2656 DALGRAYVREGGFLHDAAEFDAEFFGVSPREALAMDPQQRMLLTTSWEAFERAGIEPASL 2715
Qy 2198 HGTPTGVFTGTNGQDYALRVHNAGQSTDGFALTGTAGSVISGRISYTFGFEGPAVSVDTA 2257
Db 2716 RGSSTGVFIGLSYQDYAARVPNAPRGVEGYLLTGSTPSVASGRIAYTFGLEGPATTVDTA 2775
Qy 2258 CSSSLVALHLACQALRAGECSMALAGGVTVMSSPGAFVEFSRQRGLAADGHCKAFSAAAD 2317
Db 2776 CSSSLTALHLAVRALRSGECTMALAGGVAMMATPHMFVEFSRQRALAPDGRSKAFSADAD 2835
Qy 2318 GTGWGEGVGMLLVERLSDAHRNGHRVLAVVRGSAVNQDGASNGLTAPNGPSQQRVIRQAL 2377
Db 2836 GFGAAEGVGLLLVERLSDARRNGHPVLAVVRGTAVNQDGASNGLTAPNGPSQQRVIRQAL 2895
Qy 2378 ANAGLSAGDVDAVEAHGTGTTLGDPIEAQALLATYGQDRAGEGPLWLGSVKSNVGHTQAA 2437
Db 2896 ADARLAPGDIDAVETHGTGTSLGDPIEAQGLQATYGKERPAERPLAIGSVKSNIGHTQAA 2955
Qy 2438 AGVAGVIKMVMALRHGLLPRTLHVDEPSPHVDWSAGAVQLLTETVPWPGGEGRLRRAGVS 2497
Db 2956 AGAAGIIKMVLAMRHGTLPKTLHADEPSPHVDWANSGLALVTEPIDWPAGTGP-RRAAVS 3014

```
Qy       2498 SFGVSGTNAHVILEEAP--------ADDVP----------GGPPAGEGDAGSDDEAAA 2537
Db       3015 SFGISGTNAHVVLEQAPDAAGEVLGADEVPEVSETVAMAGTAGTSEVAEGSEASEAPAAP 3074

Qy       2538 GS-----PGVWPWLVSAKSQPALRAQAQALHAHLTDHPGLDLA---------DVGYTLAH 2583
Db       3075 GSREASLPGHLPWVLSAKDEQSLRGQAAALHAWLSE-PAADLSDADGPARLRDVGYTLAT 3133

Qy       2584 ARAVFDHRATLIAADRDTFLQALQALAAGEPHPAVIHSSAPGGTGTGEAAGKTAFICSGQ 2643
Db       3134 SRTAFAHRAAVTAADRDGFLDGLATLAQGGTSAHVHLDTARDGT--------TAFLFTGQ 3185

Qy       2644 GTQRPGMAHGLYHTHPVFAAALNDICTHLDPHLDHPLLPLLTQNDNDNEDAAALLQQTRY 2703
Db       3186 GSQRPGAGRELYDRHPVFARALDEICAHLDGHLELPLLDVMFAAEGSAE--AALLDETRY 3243

Qy       2704 AQPALPAFQVALHRLLTDGYHITPHYYAGHSLGEITAAHLAGILTLTDATTLITQRATLM 2763
Db       3244 TQCALPALEVALFRLV-ESWGMRPAALLGHSVGEIAAAHVAGVFSLADAARLVAARGRLM 3302

Qy       2764 QTMPP-GTMTTLHTTPHHITHHLTAHE---NDLAIAAINTPTSLVISGTPHTVQHITTLC 2819
Db       3303 QELPAGGAMLAVQAAEDEIRVWLETEERYAGRLDVAAVNGPEAAVLSGDADAARERAEAYW 3362

Qy       2820 QQQGIKTKTLPTNHAFHSPHTNPILNQLHQHTQTLTYHPPHTPLITANT----PPDQLLT 2875
Db       3363 SGLGRRTRALRVSHAFHSAHMDGMLDGFRAVLETVEFRRPSLTVVSNVTGLAAGPDDLCD 3422

Qy       2876 PHYWTQQARNTVDYATTTQTLHQHGVTTYIELGPDNTLTTLTHHNLPNPPTTTLT---LT 2932
Db       3423 PEYWVRHVRGTVRFLDGVRVLRDLGVRTCLELGPDGVLTAMAADGLADTPADSAAGSPVG 3482

Qy       2933 HPHHHPQTH---------LLTNL-----AKTTTTWHPHHYTH-HDNQPHTHT-------- 2969
Db       3483 SPAGSPADSAAGALRPRPLLVALLRRKRSETETVADALGRAHAHGTGPDWHAWFAGSGAH 3542

Qy       2970 HLDLPTYPFQHHHYWLESTQPGAG---NVSAAGLDPTEHPLLGATLELATDGGALLAGRL 3026
Db       3543 RVDLPTYSFRRDRYWLDA--PAADTAVDTAGLGLGTADHPLLGAVVSLPDRDGLLLTGRL 3600

Qy       3027 SLRSHPWLADHAVGGTVLLSGATFLELALHAGTYVGCDRVDELTLHAPLVVPVDGGVSVQ 3086
Db       3601 SLRTHPWLADHAVLGSVLLPGAAMVELAAHAAESAGLRDVRELTLLEPLVLPEHGGVELR 3660

Qy       3087 VGVAAADGE--------GRRLVSVYARGGSACGGGGASGGVWTCHASGVLVEAAAGGVVV 3138
Db       3661 VTVGAPAGEPGGESAGDGARPVSLHSRLADA-----PAGTAWSCHATGLLATDRPELPVA 3715

Qy       3139 DGLAGVWPPRGAVAVDVDGVRDRLAGAGCVLGPVFSGLRAVWRDGGDLLAEVCLP----- 3193
Db       3716 PDRAAMWPPQGAEEVPLDGLYERLDGNGLAFGPLFQGLNAVWRYEGEVFADIALPATTNA 3775

Qy       3194 -----EEAWGDAAG--FGLHPALLDGVVQPLSVLLPGGTGFGEGAGFGEGVRVPAVWGGV 3246
Db       3776 TAPATANGGGSAAAAPYGIHPALLDASLHAIAV---GGL-----VDEPELVRVPFHWSGV 3827

Qy       3247 SLHRAGVTGVRVRVSAVGRGGGREAVSVVVGDEAGVPVASVDRLELRPVDMGQLRAVSVS 3306
Db       3828 TVHAAGAAAARVRLAS----AGTDAVSLSLTDGEGRPLVSVERLTLRPVTADQAAASRVG 3883

Qy       3307 AGRRGSLYAVQW---------------AEVGPVPVCGQ--------------------A 3330
Db       3884 ----GLMHRVAWRPYALASSGEQDPHATSYGPTAVLGKDELKVAAALESAGVEVGLYPDL 3939

Qy       3331 WAWHEDVGESGGGPVPGVVVLRCPDAGAGGGGGGGGGGGVGEVVGGVLGVVQGWLGLERF 3390
Db       3940 AALSQDV--AAGAPAPRTVLAPLP-----AGPADGGAEGVRGTVARTLELLQAWLADEHL 3992

Qy       3391 AGSRLVVVTRGAV----VAGPEDGPVDVVGASVWGLVRSAQAEHPDRFVLLDLDTDTGT- 3445
Db       3993 AGTRLLLVTRGAVRDPEGSGADDGGEDLSHAAAWGLVRTAQTENPGRFGLLDLADDASSY 4052

Qy       3446 ----DLDTGAGAGWGVDGGRVAAVVACGEPQLAVRGERLLAARLKRLESSGDVPAQRSGD 3501
Db       4053 RTLPSVLSDAGLR--------------DEPQLALHDGTIRLARL---------------- 4082

Qy       3502 TRARRSDVPAQRSGGVPARRSVDVSGREVLPWLSGGSVLVTGGTGVLGAAVARHLAGVCG 3561
Db       4083 -----ASVRPETGTAAPA----------LAP---EGTVLLTGGTGGLGGLVARHVVGEWG 4124

Qy       3562 VRDLLLVSRRGPDAPGAEGLRAELAALGAEVRIVACDVGERREVVRLLEGVPAGCPLTGV 3621
Db       4125 VRRLLLVSRRGTDAPGADELVHELEALGADVSVAACDVADREALTAVLDAIPAEHPLTAV 4184

Qy       3622 VHAAGVLDDATIASLTPERLGTVFAAKVDAALLLDELTR--GMELSAFVLFSSAAGILGS 3679
Db       4185 VHTAGVLSDGTLPSMTTEDVEHVLRPKVDAAFLLDELTSTPAYDLAAFVMFSSAAAVFGG 4244

Qy       3680 AGQGNYAAANAALDALAYRRRAAGLPGVSLAWGLWEEASGMTGHLAGTDHRRIIRSGLHP 3739
Db       4245 AGQGAYAAANATLDALAWRRRAAGLPALSLGWGLWAETSGMTGELGQADLRRMSRAGIGG 4304

Qy       3740 MSTPDALALFDAALALDR-PVLLP-----ADLRPA-----PPLPPLLQDLLPATRRR--- 3785
Db       4305 ISDAEGIALLDAALRDDRHPVLLPLRLDAAGLRDAAGNDPAGIPALFRDVVGARTVRARP 4364

Qy       3786 ------TTRTT--TTGGADNGAQLHA-----RLAGQTHEQQHTTLLALVRSHIATVLGHT 3832
Db       4365 SAASASTTAGTAGTPGTADGAAETAAVTLADRAATVDGPARQRLLLEFVVGEVAEVLGHA 4424

Qy       3833 TPDTIPPDRAFRDLGFDSLTAVELRNRLSRTTGLRLPTTLAFDHPNPTTLTHHLHTQL-- 3890
Db       4425 RGHRIDAERGFLDLGFDSLTAVELRNRLNSAGGLALPATLVFDHPSPAALASHLDAELPR 4484

Qy       3891 -QPQPDNA-----------------VAPVLAELDKLESALSALDKTDS-ASERVTLRLK 3930
Db       4485 GASDQDGAGNRNGNENGTTASRSTAETDALLAQLTRLEGALVLTGLSDAPGSEEVLEHLR 4544

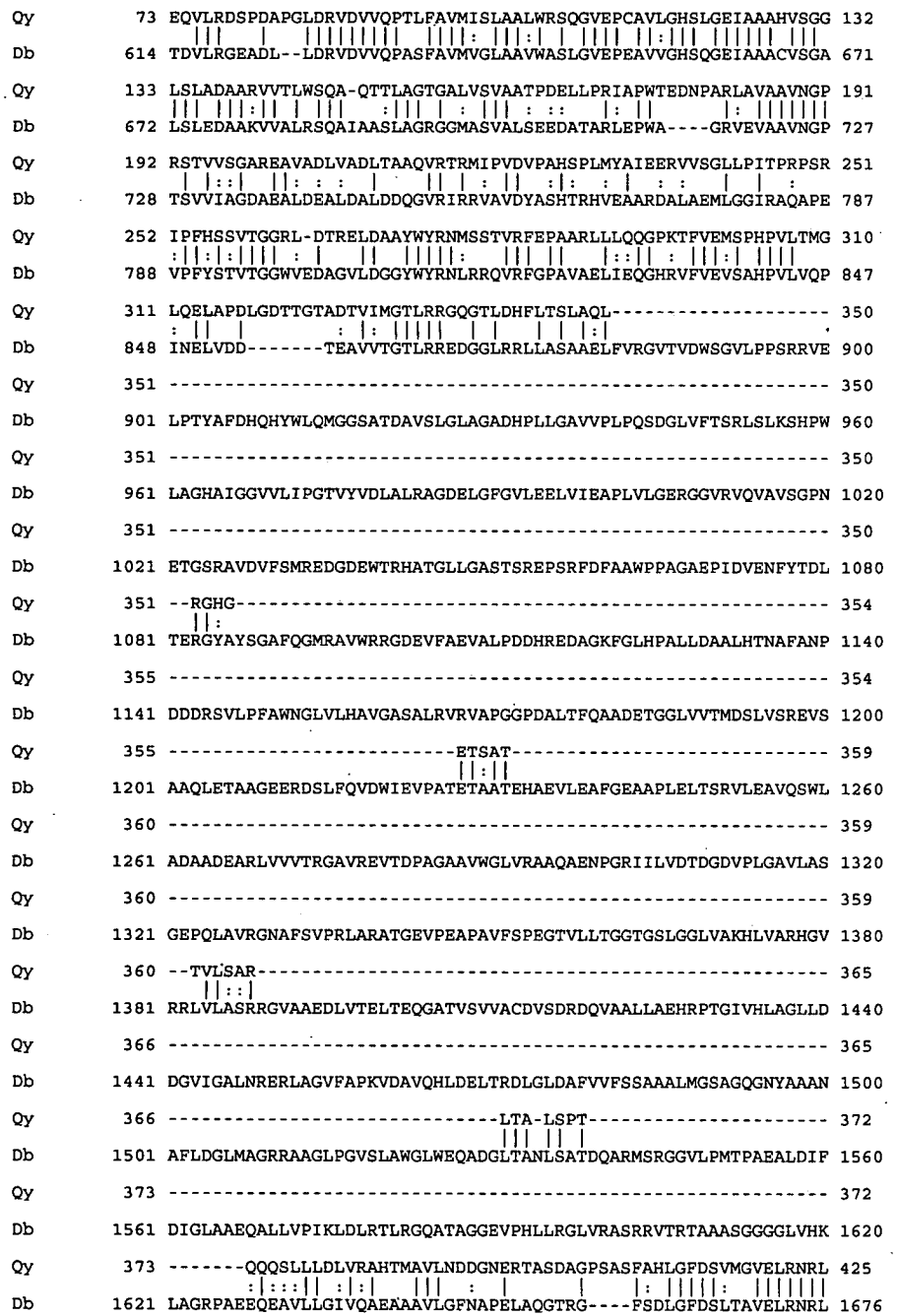
Qy       3931 SLMLR------------------------WNAPQHPTAESADDD--EKFTSATEAEIFK 3963
Db       4545 SLRSMVTGETGTGTASGAPDGAGSGAEDRPWAAGDGAGGGSEDGAGVPDFMNASAEELFG 4604

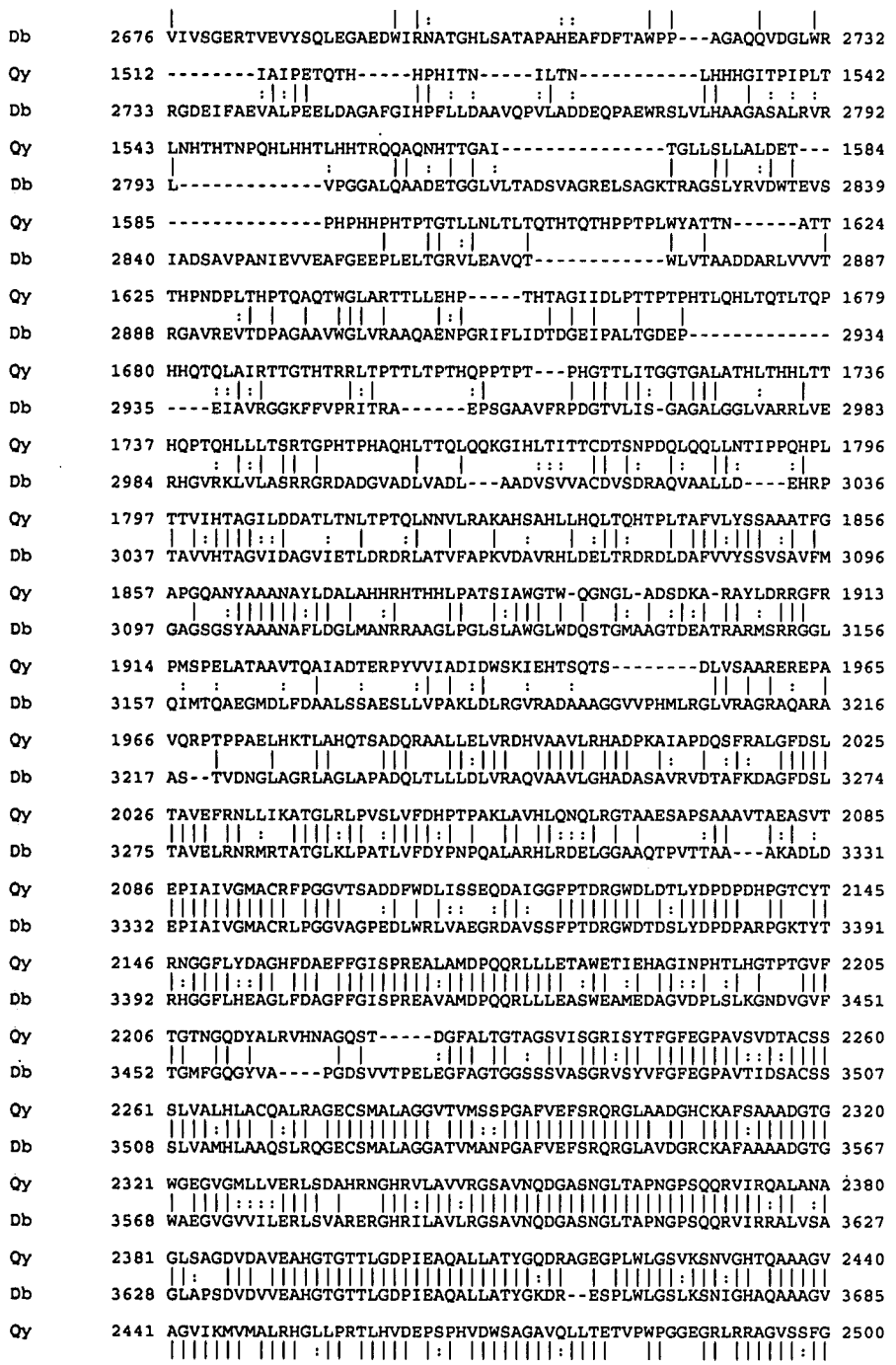
Qy       3964 FIDND 3968
Db       4605 LLDQD 4609
```

RESULT 2
T17464
rifamycin polyketide synthase modules 4-6 - Amycolatopsis mediterranei
C;Species: Amycolatopsis mediterranei
C;Date: 02-Sep-2000 #sequence_revision 02-Sep-2000 #text_change 17-Nov-2000
C;Accession: T17464
R;Schupp, T.
submitted to the EMBL Data Library, December 1997
A;Reference number: Z18802
A;Accession: T17464
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-5069 <SCH>
A;Cross-references: EMBL:AJ223012; NID:e1227119; PID:e1227121; PIDN:CAA11036.1
A;Experimental source: strain LBG A3136
C;Superfamily: acyl carrier protein homology
C;Keywords: carrier protein
F;1631-1702/Domain: acyl carrier protein homology <ACP1>
F;3238-3309/Domain: acyl carrier protein homology <ACP2>
F;4939-5010/Domain: acyl carrier protein homology <ACP3>

 Query Match 34.1%; Score 7098; DB 2; Length 5069;
 Best Local Similarity 37.3%; Pred. No. 8.4e-315;
 Matches 1791; Conservative 467; Mismatches 1346; Indels 1202; Gaps 100;

```
Qy         13 AAGEVLGVADEADGGVVFVFPGQGPQWPGMGRELLDASDVFRESVRACEAAFAPYVDWSV 72
Db        554 APGLLSGRGSGVPGKVVWVFPGQGTQWAGMGRELLDSSEVFAARIAECETALGRWVDWSL 613
```


Qy 73 EQVLRDSPDAPGLDRVDVVQPTLFAVMISLAALWRSQGVEPCAVLGHSLGEIAAAHVSGG 132
Db 614 TDVLRGEADL--LDRVDVVQPASFAVMVGLAAVWASLGVEPEAVVGHSQGEIAAACVSGA 671
Qy 133 LSLADAARVVTLWSQA-QTTLAGTGALVSVAATPDELLPRIAPWTEDNPARLAVAAVNGP 191
Db 672 LSLEDAAKVVALRSQAIAASLAGRGGMASVALSEEDATARLEPWA----GRVEVAAVNGP 727
Qy 192 RSTVVSGAREAVADLVADLTAAQVRTRMIPVDVPAHSPLMYAIEERVVSGLLPITPRPSR 251
Db 728 TSVVIAGDAEALDEALDALDDQGVRIRRVAVDYASHTRHVEAARDALAEMLGGIRAQAPE 787
Qy 252 IPFHSSVTGGRL-DTRELDAAYWYRNMSSTVRFEPAARLLLQQGPKTFVEMSPHPVLTMG 310
Db 788 VPFYSTVTGGWVEDAGVLDGGYWYRNLRRQVRFGPAVAELIEQGHRVFVEVSAHPVLVQP 847
Qy 311 LQELAPDLGDTTGTADTVIMGTLRRGQGTLDHFLTSLAQL---------- 350
Db 848 INELVDD-------TEAVVTGTLRREDGGLRRLLASAAELFVRGVTVDWSGVLPPSRRVE 900
Qy 351 ---------- 350
Db 901 LPTYAFDHQHYWLQMGGSATDAVSLGLAGADHPLLGAVVPLPQSDGLVFTSRLSLKSHPW 960
Qy 351 ---------- 350
Db 961 LAGHAIGGVVLIPGTVYVDLALRAGDELGFGVLEELVIEAPLVLGERGGVRVQVAVSGPN 1020
Qy 351 ---------- 350
Db 1021 ETGSRAVDVFSMREDGDEWTRHATGLLGASTSREPSRFDFAAWPPAGAEPIDVENFYTDL 1080
Qy 351 --RGHG---------- 354
Db 1081 TERGYAYSGAFQGMRAVWRRGDEVFAEVALPDDHREDAGKFGLHPALLDAALHTNAFANP 1140
Qy 355 ---------- 354
Db 1141 DDDRSVLPFAWNGLVLHAVGASALRVRVAPGGPDALTFQAADETGGLVVTMDSLVSREVS 1200
Qy 355 ----------ETSAT---------- 359
Db 1201 AAQLETAAGEERDSLFQVDWIEVPATETAATEHAEVLEAFGEAAPLELTSRVLEAVQSWL 1260
Qy 360 ---------- 359
Db 1261 ADAADEARLVVVTRGAVREVTDPAGAAVWGLVRAAQAENPGRIILVDTDGDVPLGAVLAS 1320
Qy 360 ---------- 359
Db 1321 GEPQLAVRGNAFSVPRLARATGEVPEAPAVFSPEGTVLLTGGTGSLGGLVAKHLVARHGV 1380
Qy 360 --TVLSAR---------- 365
Db 1381 RRLVLASRRGVAAEDLVTELTEQGATVSVVACDVSDRDQVAALLAEHRPTGIVHLAGLLD 1440
Qy 366 ---------- 365
Db 1441 DGVIGALNRERLAGVFAPKVDAVQHLDELTRDLGLDAFVVFSSAAALMGSAGQGNYAAAN 1500
Qy 366 ----------LTA-LSPT---------- 372
Db 1501 AFLDGLMAGRRAAGLPGVSLAWGLWEQADGLTANLSATDQARMSRGGVLPMTPAEALDIF 1560
Qy 373 ---------- 372
Db 1561 DIGLAAEQALLVPIKLDLRTLRGQATAGGEVPHLLRGLVRASRRVTRTAAASGGGGLVHK 1620
Qy 373 -------QQQSLLLDLVRAHTMAVLNDDGNERTASDAGPSASFAHLGFDSVMGVELRNRL 425
Db 1621 LAGRPAEEQEAVLLGIVQAEAAAVLGFNAPELAQGTRG----FSDLGFDSLTAVELRNRL 1676
Qy 426 SKATGLRLPVTLIFDHTTPAAVAARLRTAALGHLDEDTAPVPDSPSGHGGTAAAD--DPI 483
Db 1677 SAATGVKLPATLVFDYPTPVALARHLRE----ELGETVAGAPATPV----TTVADAGEPI 1728
Qy 484 AIIGMACRFPGGVRSPKDLWELAASGGDAIGPFPTDRGWPTEQRHAQDPTQPGTFYPQGG 543
Db 1729 AIVGMACRLPGGVMSPDDLWRMVAEGRDGMSPFPGDRGWDLDGLFDSDPERPGTAYIRQG 1788
Qy 544 GFLHDAAHFDAGFFGISPREALAMDPQQRLLLETSWEAFERAGIDPLSVRGSRTGVFAGA 603
Db 1789 GFLHEAALFDPGFFGISPREALAMDPQQRLLLEASWEALERAGIDPTKARGDAVGVFSGV 1848
Qy 604 LSFDYGPRMDTASSEGAADVEGHILTGTTGSVLSGRIAYSFGLEGPAITVDTGCSASLVT 663
Db 1849 SIHDYLESL----SNMPAELEGFVTTATAGSVASGRVSYTFGFEGPAVTVDTACSSSLVA 1904
Qy 664 LHLACQSLRSGECTLALAGGVSVM-STLGMFIEFSRQRGLSVDGRCKAYSAAADGTGWGE 722
Db 1905 IHLAAQALRQGECTMALAGGVAVMGSPIGV-IGMSRQRGMAEDGRVKAFADGADGTVLSE 1963
Qy 723 GVGMLLVERLSDAVRLGHRVLAVVRGSAVNQDGASNGLTAPNGPAQERVIRQALANAGLS 782
Db 1964 GVGIVVLERLSVARERGHRVLAVLRGSAVNQDGASNGLTAPNGPSQQRVIRSALAGAGLQ 2023
Qy 783 VADVDVVEGHGTGTTLGDPIEAQALLATYGQRAGDRPLWLGSLKSNIGHTMAAAGVGGVI 842
Db 2024 PSEVDVVEAHGTGTALGEPIEAQALLATYG-KSRETPLWLGSLKSNIGHTQAAAGVAAVI 2082
Qy 843 KMVMALREGVLPRTLHVDKPSPQVDWSAGAVRLLTEAVPWPGDAAGRLRRAGVSSFGIGG 902
Db 2083 KMVQALRQDTLPPTLHVQEPTKQVDWSAGAVELLTEGREWARN--GHPRRAGVSSFGISG 2140
Qy 903 TNAHVILEEAPAAGGCVAGGGVLEGAPGLAISVAESVAAPVAVSAPVAESVPVPVPVPVP 962
Db 2141 TNAHLILEEAPADD----------TAEADVPDAV-------VP 2166
Qy 963 VPVSARSEAGLRAQAEALRQYVAVRPDVSLADVGAGLACGRAVLEHRAVVLAADREELVQ 1022
Db 2167 VVISARSTGSLAGQAGRLAAF--LDGDVPLTRVAGALLSTRATLTDRAVVVAGSAEEARA 2224
Qy 1023 GLGALAAGEPDRRVTTGHAPGGDRGGVVFVFPGQGGQWAGMGVRLLASSPVFARRMQACE 1082
Db 2225 GLTALARGESASGLVTGTA--GMPGKTVWVFPGQGTQWAGMGRELLEASPVFAERIEECA 2282
Qy 1083 EALAPWVDWSVVDILRRDAGDAVWERADVVQPVLFSVMVSLAALWRSYGIEPDAVLGHSQ 1142
Db 2283 AALQPWIDWSLLDVLR---GEGELDRVDVLQPACFAVMVGLAAVWASVGVVPDAVLGHSQ 2339
Qy 1143 GEIAAAHVCGALSLKDAAKTVALRSRALAA-VRGRGGMASVPLPAQEVEQLIGERWAGRL 1201
Db 2340 GEIAAACVSGALSLEDAAKVVALRSQAIAAELSGRGGMASIQLSHDEVAARLAP-WAGRV 2398
Qy 1202 WVAAVNGPRSTAVSGDAEAVDEVLAYCAGTGVRARRIPVDYASHCPHVQPLREELLELLG 1261
Db 2399 EIAAVNGPASVVIAGDAEALTEAVEVLGG-----RRVAVDYASHTRHVEDIQDTLAETLA 2453
Qy 1262 DISPQPSGVPFFSTVEGTWL-DTTTLDAAYWYRNLHQPVRFSDAVQALADDGHRVFVEVS 1320
Db 2454 GIDAQAPVVPFYSTVAGEWITDAGVVDGGYWYRNLRNQVGFGPAVAELIEQGHGVFVEVS 2513
Qy 1321 PHPTLVPAIEDTTEDTAEDVTAIGSLRRGDNDTRRFLTALAHTHTTGIGTPTTWHHHYTH 1380
Db 2514 AHPVLVQPISELT-----DAVVTGTLRRDDGGVRRLLTSMAELFVRGV--PVDW------ 2560
Qy 1381 HHTHPHPHTHLDLPTYPFQHQHYWLESSQPGAGSGSGAGAGSGAGSGRAGT--------- 1431
Db 2561 --ATMAPPARVELPTYAFDHQHFWL--SPPAVADAPALGL-AGADHPLLGAVLPLPQSDG 2615
Qy 1432 ----------AGGTAEVESRFWDAVARQDLE--TVATTLAVPPSAGLDTV 1469
Db 2616 LVFTSRLSVRTHPWLADGVPAAALVELAVRAGDEAGCPVLADLTVEKLLVLPESGGLRVQ 2675
Qy 1470 V----------PALSAWHRHQHDQARINTWTYQ----ETWKPLTLPTTHQPHQTWL 1511


Db 2676 VIVSGERTVEVYSQLEGAEDWIRNATGHLSATAPAHEAFDFTAWPP---AGAQQVDGLWR 2732
Qy 1512 --------IAIPETQTH-----HPHITN-----ILTN----------LHHHGITPIPLT 1542
Db 2733 RGDEIFAEVALPEELDAGAFGIHPFLLDAAVQPVLADDEQPAEWRSLVLHAAGASALRVR 2792
Qy 1543 LNHTHTNPQHLHHTLHHTRQQAQNHTTGAI--------------TGLLSLLALDET--- 1584
Db 2793 L-------------VPGGALQAADETGGLVLTADSVAGRELSAGKTRAGSLYRVDWTEVS 2839
Qy 1585 -------------PHPHHPHTPTGTLLNLTLTQTHTQTHPPTPLWYATTN------ATT 1624
Db 2840 IADSAVPANIEVVEAFGEEPLELTGRVLEAVQT-----------WLVTAADDARLVVVT 2887
Qy 1625 THPNDPLTHPTQAQTWGLARTTLLEHP-----THTAGIIDLPTTPTPHTLQHLTQTLTQP 1679
Db 2888 RGAVREVTDPAGAAVWGLVRAAQAENPGRIFLIDTDGEIPALTGDEP------------ 2934
Qy 1680 HHQTQLAIRTTGTHTRRLTPTTLTPTHQPPTPT---PHGTTLITGGTGALATHLTHHLTT 1736
Db 2935 ----EIAVRGGKFFVPRITRA------EPSGAAVFRPDGTVLIS-GAGALGGLVARRLVE 2983
Qy 1737 HQPTQHLLLTSRTGPHTPHAQHLTTQLQQKGIHLTITTCDTSNPDQLQQLLNTIPPQHPL 1796
Db 2984 RHGVRKLVLASRRGRDADGVADLVADL---AADVSVVACDVSDRAQVAALLD----EHRP 3036
Qy 1797 TTVIHTAGILDDATLTNLTPTQLNNVLRAKAHSAHLLHQLTQHTPLTAFVLYSSAAATFG 1856
Db 3037 TAVVHTAGVIDAGVIETLDRDRLATVFAPKVDAVRHLDELTRDRDLDAFVVYSSVSAVFM 3096
Qy 1857 APGQANYAAANAYLDALAHHRHTHHLPATSIAWGTW-QGNGL-ADSDKA-RAYLDRRGFR 1913
Db 3097 GAGSGSYAAANAFLDGLMANRRAAGLPGLSLAWGLWDQSTGMAAGTDEATRARMSRRGGL 3156
Qy 1914 PMSPELATAAVTQAIADTERPYVVIADIDWSKIEHTSQTS--------DLVSAARE-REPA 1965
Db 3157 QIMTQAEGMDLFDAALSSAESLLVPAKLDLRGVRADAAAGGVVPHMLRGLVRAGRAQARA 3216
Qy 1966 VQRPTPPAELHKTLAHQTSADQRAALLELVRDHVAAVLRHADPKAIAPDQSFRALGFDSL 2025
Db 3217 AS--TVDNGLAGRLAGLAPADQLTLLLDLVRAQVAAVLGHADASAVRVDTAFKDAGFDSL 3274
Qy 2026 TAVEFRNLLIKATGLRLPVSLVFDHPTPAKLAVHLQNQLRGTAAESAPSAAAVTAEASVT 2085
Db 3275 TAVELRNRMRTATGLKLPATLVFDYPNPQALARHLRDELGGAAQTPVTTAA---AKADLD 3331
Qy 2086 EPIAIVGMACRFPGGVTSADDFWDLISSEQDAIGGFPTDRGWDLDTLYDPDPDHPGTCYT 2145
Db 3332 EPIAIVGMACRLPGGVAGPEDLWRLVAEGRDAVSSFPTDRGWDTDSLYDPDPARPGKTYT 3391
Qy 2146 RNGGFLYDAGHFDAEFFGISPREALAMDPQQRLLLETAWETIEHAGINPHTLHGTPTGVF 2205
Db 3392 RHGGFLHEAGLFDAGFFGISPREAVAMDPQQRLLLEASWEAMEDAGVDPLSLKGNDVGVF 3451
Qy 2206 TGTNGQDYALRVHNAGQST-----DGFALTGTAGSVISGRISYTFGFEGPAVSVDTACSS 2260
Db 3452 TGMFGQGYVA----PGDSVVTPELEGFAGTGGSSSVASGRVSYVFGFEGPAVTIDSACSS 3507
Qy 2261 SLVALHLACQALRAGECSMALAGGVTVMSSPGAFVEFSRQRGLAADGHCKAFSAAADGTG 2320
Db 3508 SLVAMHLAAQSLRQGECSMALAGGATVMANPGAFVEFSRQRGLAVDGRCKAFAAAADGTG 3567
Qy 2321 WGEGVGMLLVERLSDAHRNGHRVLAVVRGSAVNQDGASNGLTAPNGPSQQRVIRQALANA 2380
Db 3568 WAEGVGVVILERLSVARERGHRILAVLRGSAVNQDGASNGLTAPNGPSQQRVIRRALVSA 3627
Qy 2381 GLSAGDVDAVEAHGTGTTLGDPIEAQALLATYGQDRAGEGPLWLGSVKSNVGHTQAAAGV 2440
Db 3628 GLAPSDVDVVEAHGTGTTLGDPIEAQALLATYGKDR--ESPLWLGSLKSNIGHAQAAAGV 3685
Qy 2441 AGVIKMVMALRHGLLPRTLHVDEPSPHVDWSAGAVQLLTETVPWPGGEGRLRRAGVSSFG 2500
Db 3686 AGVIKMVQALRHEVLPPTLHVDRPTPEVDWSAGAVELLTEAREWP-RNGRPRRAGVSAFG 3744
Qy 2501 VSGTNAHVILEEAPADD-VPGGPPAGEGDAGSDDEAAAGSPGV--WPWLVSAKSQPALRA 2557
Db 3745 VSGTNAHLILEEAPAEEPVP----------TPEVPLVPVVVSARSRASLAG 3785
Qy 2558 QAQALHAHLTDHPGLDLADVGYTLAHARAVFDHRATLIAADRDTFLQALQALAAGEPHPA 2617
Db 3786 QAGRLAGFVAG--DASLAGVARALVTNRAALTERAVMVVGSREEAVTNLEALARGEDPAA 3843
Qy 2618 VIHSSAPGGTGTGEAAGKTAFICSGQGTQRPGMAHGLYHTHPVFAAALNDICTHLDPHLD 2677
Db 3844 VV-------TGRAGSPGKLVWVFPGQGSQWIGMGRELLDSSPVFAERVAECAAALEPWID 3896
Qy 2678 HPLLPLLTQNDNDNEDAAALLQQTRYAQPALFAFQVALHRLLTDGYHITPHYYAGHSLGE 2737
Db 3897 WSLLDVL-RGESD------LLDRVDVVQPASPAMMVGL-AAVWQSVGVRPDAVVGHSQGE 3948
Qy 2738 ITAAHLAGILTLTDATTLITQRATLMQTMPPGT--MTTLHTTPHHITHHLTAHENDLAIA 2795
Db 3949 IAAACVSGALSLQDAAKVVALRSQAIATRLAGRGGMASVALSEEDATAWLAPWADRVQVA 4008
Qy 2796 AINTPTSLVISGTPHTVQHITTLCQQQGIKTKTLPTNHAFHSPHTNPILNQLHQHTQTLT 2855
Db 4009 AVNSPASVVIAGEAQALDEVVDALSGQEVRVRRVAVDYGSHTNQVEAIEDLLAETLAGIE 4068
Qy 2856 YHPPHTP----LITANTPPDQLLTPHYWTQQARNTVDYATTTQTLHQHGVTTYIELGPDN 2911
Db 4069 AQAPKVPFYSTLIGDWIRDAGIVDGGYWYRNLRNQVGFGPAVAELVRQGHGVFVEVSAHP 4128
Qy 2912 TLTTLTHHNLPNPPTTTLTLTHPHHHPQTHLLTNLAKTTTTWHPHHYTHHDNQPHTHTHL 2971
Db 4129 VLVQ-PLSELSDDAVVTGSLRR-EDGGLRRLLTSMAELYVQGVPLDWT--AVLPRT-GRV 4183
Qy 2972 DLPTYPFQHHHYWLESTQPGAGNVSAAGLDPTEHPLLGATLELATDGGALLAGRLSLRSH 3031
Db 4184 DLPKYAFDHRHYWLRPAE-SATDAASLGQAAADHPLLGAVVELPQSDGLVFTSRLSVRTH 4242
Qy 3032 PWLADHAVGGTVLLSGATFLELALHAGTYVGCDRVDELTLHAPLVVPVDGGVSVQVGVAA 3091
Db 4243 PWLADHAVGGVVILPGSGLAELAVRAGDEAGCTALDELIIEAPLVVPAQGAVRVQVALSG 4302
Qy 3092 ADGEGRRLVSVYARGGSACGGGGASGGVWTCHASGVLVEAAAGGVVVDGLAGVWPPRGAV 3151
Db 4303 PDETGSRTVDLYSQ-----RDGGA--GTWTRHATGVLSTAPAQEPEFDFHA--WPPADAE 4353
Qy 3152 AVDVDGVRDRLAGAGCVLGPVFSGLRAVWRDGGDLLAEVCLPEEAWGDAAGFGLHPALLD 3211
Db 4354 RIDVETFYTDLAERGYGYGPAFQGLQAVWRRDGDVFAEVALPEDLRKDAGRFGVHPALLD 4413
Qy 3212 GVVQPLSVLLPGGTGFGEGAGFGEGVRVPAVWGGVSLHRAGVTGVRVRVSAVGRGGGREA 3271
Db 4414 AALQAATAV--GGDEPGQPV-------LAFAWNGLVLHAAGASALRVRLAP----SGPDT 4460
Qy 3272 VSVVVGDEAGVPVASVDRLELRPVDMGQLRAVSVSAGRRGSLYAVQWAEVGPVPVCGQAW 3331
Db 4461 LSVAAADETGGLVLTMESLVSRPVSAEQLGA-AADAG-HDAMFRVDWTELPAVP-RAELP 4517
Qy 3332 AW-----HEDVGE-SGGGPVPGVVVLRCPDAGAGGGGGGGGGGGVGEVVGGVLGVVQGWL 3385
Db 4518 PWVRIDTADDVAALAEKADAPPVVVWE----------AAGGDPALAVSSRVLEIMQAWL 4566
Qy 3386 GLERFAGSRLVVVTRGAVVAGPEDGPVDVVGASVWGLVRSAQAEHPDRFVLLDLDTDTGT 3445
Db 4567 AAPAFEEARLVVTTRGAVPAGGDHTLTDPAAAAVWGLVRSAQAEHPDRVVLLDTDGEV-- 4624
Qy 3446 DLDTGAGAGWGVDGGRVAAVVACGEPQLAVRGERLLAARLKRLESSGDVPAQRSGDTRAR 3505
Db 4625 --------------PLGAVLASGEPQLAVRGTTFFVPRLARATRLSDAPPAFDPD---- 4665
Qy 3506 RSDVPAQRSGGVPARRSVDVSGREVLPWLSGGSVLVTGGTGVLGAAVARHLAGVCGVRDL 3565
Db 4666 ----------------------------GTVLVS-GAGSLGTLVARHLVTRHGVRRV 4693

```
Qy       3566 LLVSRRGPDAPGAEGLRAELAALGAEVRIVACDVGERREVVRLLEGVPAGCPLTGVVHAA 3625
Db       4694 VLASRQGRDAEGAQDLITELTGEGADVSFVACDVSDRDQVAALLAGLP---DLTGVVHTA 4750

Qy       3626 GVLDDATIASLTPERLGTVFAAKVDAALLLDELTRGMELSAFVLFSSAAGILGSAGQGNY 3685
Db       4751 GVFEDGVIEALTPDQLANVYAAKVTAAMHLDELTRDRDLGAFVVFSSVAGVMGGGGQGPY 4810

Qy       3686 AAANAALDALAYRRRAAGLPGVSLAWGLWEEASGMTGHLAGTDHRRIIRSGLHPMSTPDA 3745
Db       4811 AAANAFLDAAMASRQAAGLPGLSLAWGLWERSSGMAAHLSEVDHARASRNGVLELTRAEG 4870

Qy       3746 LALFDAALALDRPVLLPADL----RPAPPLPPLLQDLLPATRRRTTRTTTTGGADNGAQL 3801
Db       4871 LALFDLGLRMAESLLVPIKLDLAAMRASTVPVLFRGLVRPSR---TQARTASTVDRG--L 4925

Qy       3802 HARLAGQTHEQQHTTLLALVRSHIATVLGHTTPDTIPPDRAFRDLGFDSLTAVELRNRLS 3861
Db       4926 AGRLAGLPVAERAAVLVDLVRGQVAVVLGYDGPEAVRPDTAFKDTGFDSLTSVELRNRLR 4985

Qy       3862 RTTGLRLPTTLAFDHPNPTTLTHHLHTQLQPQ---PDNAVAPVLAELDKLESALSALDKT 3918
Db       4986 EATGLKLPATLVFDYPNPLAVARYLGARLVPDGTANGNGNGNGHSEDDRLRHALAAIAAE 5045

Qy       3919 DSASER 3924
Db       5046 DAGEER 5051


RESULT 3
T17410
polyketide synthase type I - Streptomyces venezuelae
C;Species: Streptomyces venezuelae
C;Date: 02-Sep-2000 #sequence_revision 02-Sep-2000 #text_change 17-Nov-2000
C;Accession: T17410
R;Xue, Y.; Zhao, L.; Liu, H.W.; Sherman, D.H.
Proc. Natl. Acad. Sci. U.S.A. 95, 12111-12116, 1998
A;Title: A gene cluster for macrolide antibiotic biosynthesis in streptomyces venezuelae
A;Reference number: Z18773; MUID:98445333; PMID:9770448
A;Accession: T17410
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-3739 <XUE>
A;Cross-references: EMBL:AF079138; NID:g3808326; PID:g3800835; PIDN:AAC69330.1
C;Genetics:
A;Gene: pikAII
C;Superfamily: acyl carrier protein homology
C;Keywords: antibiotic biosynthesis; carrier protein
F;1445-1516/Domain: acyl carrier protein homology <ACP1>
F;3570-3641/Domain: acyl carrier protein homology <ACP2>

  Query Match             33.7%;  Score 7017;  DB 2;  Length 3739;
  Best Local Similarity   42.4%;  Pred. No. 2.7e-311;
  Matches 1676;  Conservative  390;  Mismatches 1177;  Indels  712;  Gaps   82;

Qy        477 AAADDPIAIIGMACRFPGGVRSPKDLWELAASGGDAIGPFPTDRGWPTEQRHAQDPTQPG 536
Db         32 AKAGEPVAIVGMACRLPGGVASPEDLWRLVAGGEDAISEFPQDRGWDVEGLYDPNPEATG 91

Qy        537 TFYPQGGGFLHDAAHFDAGFFGISPREALAMDPQQRLLLETSWEAFERAGIDPLSVRGSR 596
Db         92 KSYAREAGFLYEAGEFDADFFGISPREALAMDPQQRLLLEASWEAFEHAGIPAATARGTS 151

Qy        597 TGVFAGALSFDYGPRMDTASSEGAADVEGHILTGTTGSVLSGRIAYSFGLEGPAITVDTG 656
Db        152 VGVFTGVMYHDYATRL-TDVPEG---IEGYLGTGNSGSVASGRVAYTLGLEGPAVTVDTA 207

Qy        657 CSASLVTLHLACQSLRSGECTLALAGGVSVMSTLGMFIEFSRQRGLSVDGRCKAYSAAAD 716
Db        208 CSSSLVALHLAVQALRKGEVDMALAGGVTVMSTPSTFVEFSRQRGLAPDGRSKSFSSTAD 267

Qy        717 GTGWGEGVGMLLVERLSDAVRLGHRVLAVVRGSAVNQDGASNGLTAPNGPAQERVIRQAL 776
```

```
Db        268 GTSWSEGVGVLLVERLSDARRKGHRILAVVRGTAVNQDGASSGLTAPNGPSQQRVIRRAL 327

Qy        777 ANAGLSVADVDVVEGHGTGTTLGDPIEAQALLATYGQ-RAGDRPLWLGSLKSNIGHTMAA 835
Db        328 ADARLTTSDVDVVEAHGTGTRLGDPIEAQAVIATYGQGRDGEQPLRLGSLKSNIGHTQAA 387

Qy        836 AGVGGVIKMVMALREGVLPRTLHVDKPSPQVDWSAGAVRLLTEAVPWPGDAAGRLRRAGV 895
Db        388 AGVSGVIKMVQAMRHGVLPKTLHVEKPTDQVDWSAGAVELLTEAMDWPDKGDGGLRRAAV 447

Qy        896 SSFGIGGTNAHVILEEAPAAGGCVAGGGVLEGAPGLAISVAESVAAPVAVSAPVAESVPV 955
Db        448 SSFGVSGTNAHVVLEEAPA----------------------------AEETPASEATPA 478

Qy        956 PVPVP----VPVPVSARSEAGLRAQAEALRQYVAVRPDVSLADVGA---GLACGRAVLEH 1008
Db        479 VEPSVGAGLVPWLVSAKTPAALDAQIGRLAAF-ASQGRTDAADPGAVARVLAGGRAEFEH 537

Qy       1009 RAVVLAADREELVQGLGA---LAAGEPDRRVTTGHAPGGDRGGVVFVFPGQGGQWAGMGV 1065
Db        538 RAVVLGTGQDDFAQALTAPEGLIRGTP----------SDVGRVAFVFPGQGTQWAGMGA 586

Qy       1066 RLLASSPVFARRMQACEEALAPWVDWSVVDILRRDAGDAVWERADVVQPVLFSVMVSLAA 1125
Db        587 ELLDVSKEFAAAMAECESALSRYVDWSLEAVVRQAPGAPTLERVDVVQPVTFAVMVSLAK 646

Qy       1126 LWRSYGIEPDAVLGHSQGEIAAAHVCGALSLKDAAKTVALRSRALAA-VRGRGGMASVPL 1184
Db        647 VWQHHGVTPQAVVGHSQGEIAAAYVAGALTLDDAARVVTLRSKSIAAHLAGKGGMISLAL 706

Qy       1185 ----PAQEVEQLIGERWAGRLWVAAVNGPRSTAVSGDAEAVDEVLAYCAGTGVRARRIPV 1240
Db        707 SEEATRQRIENLHG------LSIAAVNGPTATVVSGDPTQIQELAQACEADGVRARIIPV 760

Qy       1241 DYASHCPHVQPLREELLELLGDISPQPSGVPFFSTVEGTWLDTTTLDAAYWYRNLHQPVR 1300
Db        761 DYASHSAHVETIESELAEVLAGLSPRTPEVPFFSTLEGAWITEPVLDGTYWYRNLRHRVG 820

Qy       1301 FSDAVQALA-DDGHRVFVEVSPHPTLVPAIEDTTEDTAEDVTAIGSLRRGDNDTRRFLTA 1359
Db        821 FAPAVETLATDEGFTHFIEVSAHPVLTMTLPET-------VTGLGTLRREQGGQERLVTS 873

Qy       1360 LAHTHTTGIG------TPTTWHHHYTHHHTHPHPHTHLDLPTYPFQHQHYWLESSQPGAG 1413
Db        874 LAEAWTNGLTIDWAPVLPTATGH------------HPELPTYAFQRRHYWLHDSPAVQG 920

Qy       1414 SGSGAGAGSGAGSGRAGTAGGTAEVESRFWDAVARQDLETVATTLAVPPSAGLDTVVPAL 1473
Db        921 SVQ---------------------------------------------------------- 923

Qy       1474 SAWHRHQHDQARINTWTYQETWKPLTLPTTHQP---HQTWLIAIPETQTHHPHITNILTN 1530
Db        924 ------------DSWRYRIDWKRLAVADASERAGLSGRWLVVVPEDRS--AEAAPVLAA 968

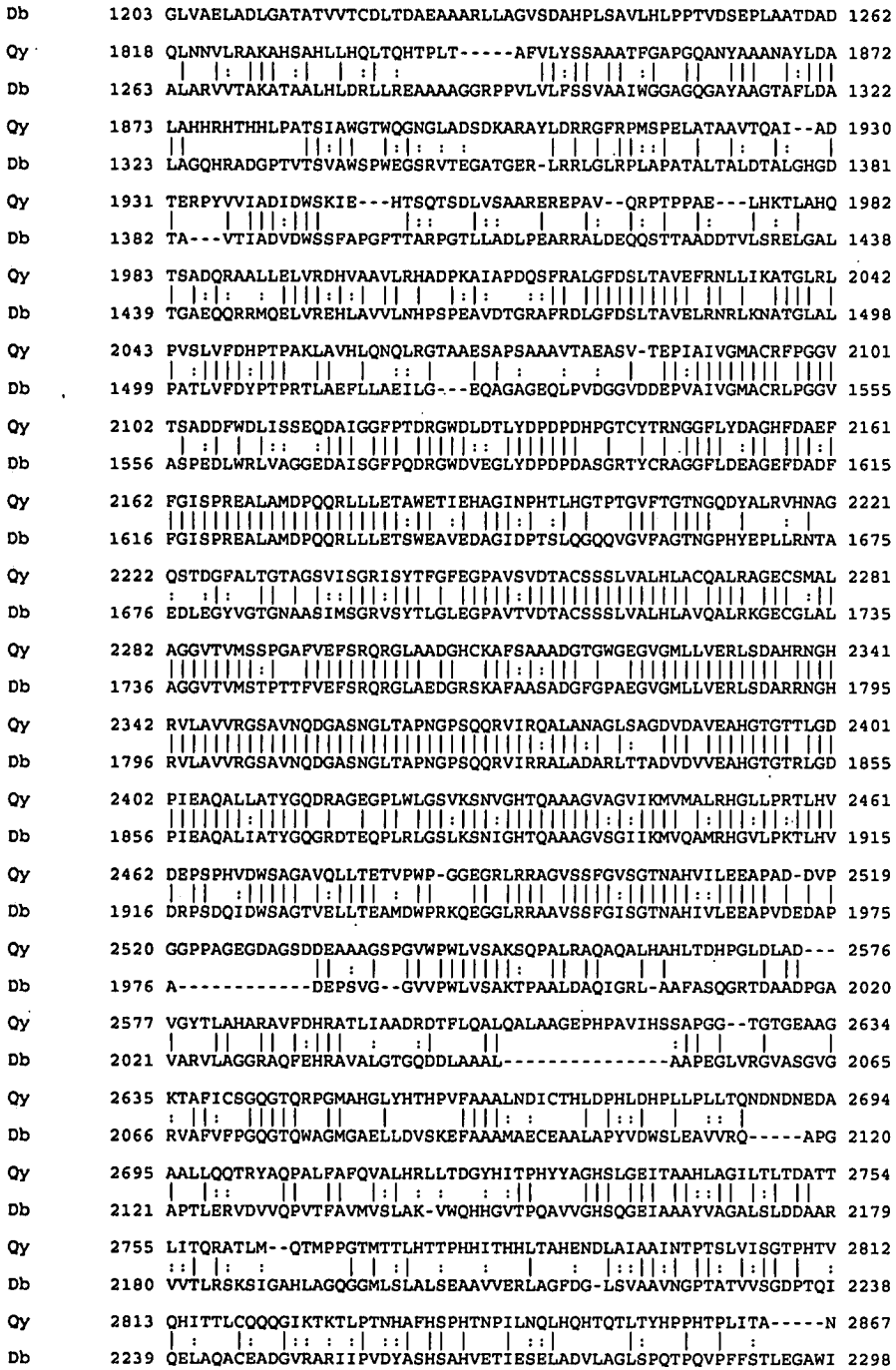
Qy       1531 LHHHGITPIPLTLNHTHTNPQHLHHTLHHTRQQAQNHTTGAITGLLSLLALDETPHPHHP 1590
Db        969 LSGAGADPVQLDV-----SPLGDRQRLAATLGEALAAAGGAVDGVLSLLAWDESAHPGHP 1023

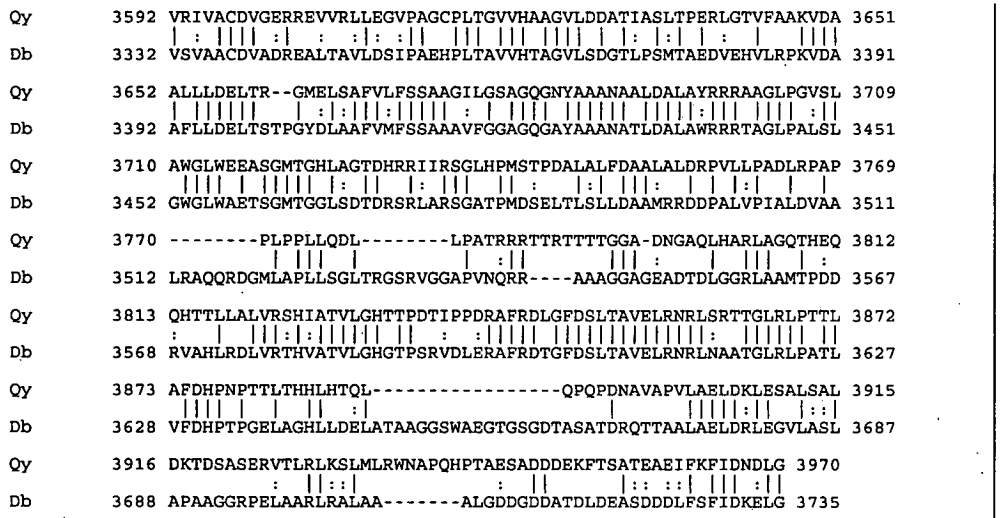
Qy       1591 HTPT-GTLLNLTLTQTHTQTHPPTPLWYATTNATTTHPNDPLTHPTQAQTWGLARTTLLE 1649
Db       1024 APFTRGTGATLTLVQALEDAGVAAPLWCVTHGAVSVGRADHVTSPAQAMVWGMGRVAALE 1083

Qy       1650 HPTHTAGIIDLPTTPTPHTLQHLTQTLTQPHHQTQLAIRTTGTHTRRLTPTTLTPTHQPP 1709
Db       1084 HPERWGGLIDLPSDADRAALDRMTTVLAGGTGEDQVAVRASGLLARRLVRASL-PAHGTA 1142

Qy       1710 TP--TPHGTTLITGGTGALATHLTHHLTTHQPTQHLLLTSRTGPH----TPHAQH----- 1758
Db       1143 SPWWQADGTVLVTGAEEPAAAEAARRLARDGAGHLLLHTTPSGSEGAEGTSGAAEDSGLA 1202

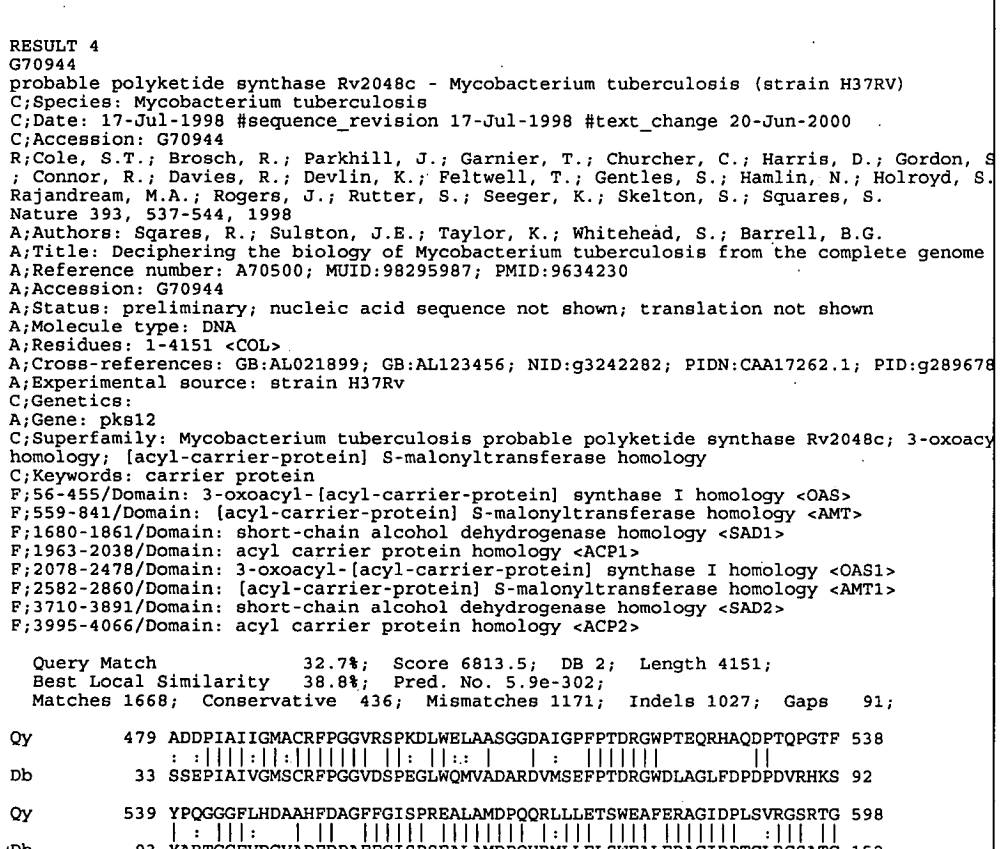
Qy       1759 -LTTQLQQKGIHLTITTCDTSNPDQLQQLLNTIPPQHPLTTVIHTAGILDDATLTNLTPT 1817
```


Db 1203 GLVAELADLGATATVVTCDLTDAEAAARLLAGVSDAHPLSAVLHLPPTVDSEPLAATDAD 1262
Qy 1818 QLNNVLRAKAHSAHLLHQLTQHTPLT-----AFVLYSSAAATFGAPGQANYAAANAYLDA 1872
Db 1263 ALARVVTAKATAALHLDRLLREAAAAGGRPPVLVLFSSVAAIWGGAGQGAYAAGTAFLDA 1322
Qy 1873 LAHHRHTHHLPATSIAWGTWQGNGLADSDKARAYLDRRGFRPMSPELATAAVTQAI--AD 1930
Db 1323 LAGQHRADGPTVTSVAWSPWEGSRVTEGATGER-LRRLGLRPLAPATALTALDTALGHGD 1381
Qy 1931 TERPYVVIADIDWSKIE---HTSQTSDLVSAARELREPAV--QRPTPPAE---LHKTLAHQ 1982
Db 1382 TA---VTIADVDWSSFAPGFTTARPGTLLADLPEARRALDEQQSTTAADDTVLSRELGAL 1438
Qy 1983 TSADQRAALLELVRDHVAAVLRHADPKAIAPDQSFRALGFDSLTAVEFRNLLIKATGLRL 2042
Db 1439 TGAEQQRRMQELVREHLAVVLNHPSPEAVDTGRAFRDLGFDSLTAVELRNRLKNATGLAL 1498
Qy 2043 PVSLVFDHPTPAKLAVHLQNQLRGTAAESAPSAAAVTAEASV-TEPIAIVGMACRFPGGV 2101
Db 1499 PATLVFDYPTPRTLAEFLLAEILG---EQAGAGEQLPVDGGVDDEPVAIVGMACRLPGGV 1555
Qy 2102 TSADDFWDLISSEQDAIGGFPTDRGWDLDTLYDPDPDHPGTCYTRNGGFLYDAGHFDAEF 2161
Db 1556 ASPEDLWRLVAGGEDAISGFPQDRGWDVEGLYDPDPDASGRTYCRAGGFLDEAGEFDADF 1615
Qy 2162 FGISPREALAMDPQQRLLLETAWETIEHAGINPHTLHGTPTGVFTGTNGQDYALRVHNAG 2221
Db 1616 FGISPREALAMDPQQRLLLETSWEAVEDAGIDPTSLQGQQVGVFAGTNGPHYEPLLRNTA 1675
Qy 2222 QSTDGFALTGTAGSVISGRISYTFGFEGPAVSVDTACSSSLVALHLACQALRAGECSMAL 2281
Db 1676 EDLEGYVGTGNAASIMSGRVSYTLGLEGPAVTVDTACSSSLVALHLAVQALRKGECGLAL 1735
Qy 2282 AGGVTVMSSPGAFVEFSRQRGLAADGHCKAFSAAADGTGWGEGVGMLLVERLSDAHRNGH 2341
Db 1736 AGGVTVMSTPTTFVEFSRQRGLAEDGRSKAFAASADGFGPAEGVGMLLVERLSDARRNGH 1795
Qy 2342 RVLAVVRGSAVNQDGASNGLTAPNGPSQQRVIRQALANAGLSAGDVDAVEAHGTGTTLGD 2401
Db 1796 RVLAVVRGSAVNQDGASNGLTAPNGPSQQRVIRRALADARLTTADVDVVEAHGTGTRLGD 1855
Qy 2402 PIEAQALLATYGQDRAGEGPLWLGSVKSNVGHTQAAAGVAGVIKMVMALRHGLLPRTLHV 2461
Db 1856 PIEAQALIATYGQGRDTEQPLRLGSLKSNIGHTQAAAGVSGIIKMVQAMRHGVLPKTLHV 1915
Qy 2462 DEPSPHVDWSAGAVQLLTETVPWP-GGEGRLRRAGVSSFGVSGTNAHVILEEAPAD-DVP 2519
Db 1916 DRPSDQIDWSAGTVELLTEAMDWPRKQEGGLRRAAVSSFGISGTNAHIVLEEAPVDEDAP 1975
Qy 2520 GGPPAGEGDAGSDDEAAAGSPGVWPWLVSAKSQPALRAQAQALHAHLTDHPGLDLAD--- 2576
Db 1976 A------------DEPSVG--GVVPWLVSAKTPAALDAQIGRL-AAFASQGRTDAADPGA 2020
Qy 2577 VGYTLAHARAVFDHRATLIAADRDTFLQALQALAAGEPHPAVIHSSAPGG--TGTGEAAG 2634
Db 2021 VARVLAGGRAQFEHRAVALGTGQDDLAAAL--------------AAPEGLVRGVASGVG 2065
Qy 2635 KTAFICSGQGTQRPGMAHGLYHTHPVFAAALNDICTHLDPHLDHPLLPLLTQNDNDNEDA 2694
Db 2066 RVAFVFPGQGTQWAGMGAELLDVSKEFAAAMAECEAALAPYVDWSLEAVVRQ-----APG 2120
Qy 2695 AALLQQTRYAQPALFAFQVALHRLLTDGYHITPHYYAGHSLGEITAAHLAGILTLTDATT 2754
Db 2121 APTLERVDVVQPVTFAVMVSLAK-VWQHHGVTPQAVVGHSQGEIAAAYVAGALSLDDAAR 2179
Qy 2755 LITQRATLM--QTMPPGTMTTLHTTPHHITHHLTAHENDLAIAAINTPTSLVISGTPHTV 2812
Db 2180 VVTLRSKSIGAHLAGQGGMLSLALSEAAVVERLAGFDG-LSVAAVNGPTATVVSGDPTQI 2238
Qy 2813 QHITTLCQQQGIKTKTLPTNHAFHSPHTNPILNQLHQHTQTLTYHPPHTPLITA-----N 2867
Db 2239 QELAQACEADGVRARIIPVDYASHSAHVETIESELADVLAGLSPQTPQVPFFSTLEGAWI 2298
Qy 2868 TPPDQLLTPHYWTQQARNTVDYATTTQTL-HQHGVTTYIELGPDNTLTTLTHHNLPNPPT 2926
Db 2299 TEP--ALDGGYWYRNLRHRVGFAPAVETLATDEGFTHFVEVSAHPVLTMA----LPETVT 2352
Qy 2927 TTLTLTHPHHHPQTHLLTNLAKT-----TTTWHPHHYTHHDNQPHTHTHLDLPTYPFQHH 2981
Db 2353 GLGTLRR-DNGGQHRLTTSLAEAWANGLTVDW-------ASLLPTTTTHPDLPTYAFQTE 2404
Qy 2982 HYWLESTQPGAGNVSAAGLDPTEHPLLGATLELATDGGALLAGRLSLRSHPWLADHAVGG 3041
Db 2405 RYWPQPDLSAAGDITSAGLGAAEHPLLGAAVALADSDGCLLTGSLSLRTHPWLADHAVAG 2464
Qy 3042 TVLLSGATFLELALHAGTYVGCDRVDELTLHAPLVVPVDGGVSVQVGVAAADGEGRRLVS 3101
Db 2465 TVLLPGTAFVELAFRAGDQVGCDLVEELTLDAPLVLPRRGAVRVQLSVGASDESGRRTFG 2524
Qy 3102 VYARGGSACGGGGASGGVWTCHASGVLVEAAAGGVVVDGLAGV-----WPPRGAVAVDVD 3156
Db 2525 LYAHPEDAPG-----EAEWTRHATGVLAARA------DRTAPVADPEAWPPPGAEPVDVD 2573
Qy 3157 GVRDRLAGAGCVLGPVFSGLRAVWRDGGDLLAEVCLPEEAWG-DAAGFGLHPALLDGVVQ 3215
Db 2574 GLYERFAANGYGYGPLFQGVRGVWRRGDEVFADVALPAEVAGAEGARFGLHPALLDAAVQ 2633
Qy 3216 PLSVLLPGGTGFGEGAGFGEGVRVPAVWGGVSLHRAGVTGVRVRVSAVGRGGGREAVSVV 3275
Db 2634 ------AAGAGRGVRRGHAAAVRLER----DLLYAVGATALRVRLAP----AGPDTVSVS 2679
Qy 3276 VGDEAGVPVASVDRLELRPVDMGQLRAVSVSAGRRGSLYAVQW--------AEVGPVPVC 3327
Db 2680 AADSSGQPVFAADSLTVLPVDPAQLAAFSDPT--LDALHLLEWTAWDGAAQALPGAVVLG 2737
Qy 3328 GQA----------WAWHEDVGES---GGGPVPGVVVLRCPDAGAGGGGGGGGGG 3368
Db 2738 GDADGLAAALRAGGTEVLSFPDLTDLVEAVDRGETPAPATVLVACPAAGP------DGPE 2791
Qy 3369 GVGEVVGGVLGVVQGWLGLERFAGSRLVVVTRGAVVAGPEDGPVDVVGASVWGLVRSAQA 3428
Db 2792 HVREALHGSLALMQAWLADERFTDGRLVLVTRDAVAARSGDGLRSTGQAAVWGLGRSAQT 2851
Qy 3429 EHPDRFVLLDL--------DTDTGTDLDTGAGAGWGVD-----GGRVAAVVACGEPQLAV 3475
Db 2852 ESPGRFVLLDLAGEARTAGDATAGDGLTTGDATVGGTSGDAALGSALATALGSGEPQLAL 2911
Qy 3476 RGERLLAARLKR----------------------------------LES----SGD- 3493
Db 2912 RDGALLVPRLARAAAPAAADGLAAADGLAALPLPAAPALWRLEPGTDGSLESLTAAPGDA 2971
Qy 3494 --VP 3495
Db 2972 ETLAPEPLGPGQVRIAIRATGLNFRDVLIALGMYPDPALMGTEGAGVVTATGPGVTHLAP 3031
Qy 3496 AQR------------------------SGDTRARRSDVPA-------------------- 3511
Db 3032 GDRVMGLLSGAYAPVVVADARTVARMPEGWTFAQGASVPVVFLTAVYALRDLADVKPGER 3091
Qy 3512 ----QRSGGVP-- 3518
Db 3092 LLVHSAAGGVGMAAVQLARHWGVEVHGTASHGKWDALRALGLDDAHIASSRTLDFESAFR 3151
Qy 3519 ---------------ARRSVDVS---------------------------------- 3526
Db 3152 AASGGAGMDVVLNSLAREFVDASLRLLGPGGRFVEMGKTDVRDAERVAADHPGVGYRAFD 3211
Qy 3527 --GREVLPWLS 3535
Db 3212 LGEAGPERIGEMLAEVIALFEDGVLRHLPVTTWDVRRARDAFRHVSQARHTGKVVLTMPS 3271
Qy 3536 G----GSVLVTGGTGVLGAAVARHLAGVCGVRDLLLVSRRGPDAPGAEGLRAELAALGAE 3591
Db 3272 GLDPEGTVLLTGGTGALGGIVARHVVGEWGVRRLLLVSRRGTDAPGAGELVHELEALGAD 3331



RESULT 4
G70944
probable polyketide synthase Rv2048c - Mycobacterium tuberculosis (strain H37RV)
C;Species: Mycobacterium tuberculosis
C;Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
C;Accession: G70944
R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S
; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A;Authors: Sqares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A;Reference number: A70500; MUID:98295987; PMID:9634230
A;Accession: G70944
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-4151 <COL>
A;Cross-references: GB:AL021899; GB:AL123456; NID:g3242282; PIDN:CAA17262.1; PID:g289678
A;Experimental source: strain H37Rv
C;Genetics:
A;Gene: pks12
C;Superfamily: Mycobacterium tuberculosis probable polyketide synthase Rv2048c; 3-oxoacy
homology; [acyl-carrier-protein] S-malonyltransferase homology
C;Keywords: carrier protein
F;56-455/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS>
F;559-841/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT>
F;1680-1861/Domain: short-chain alcohol dehydrogenase homology <SAD1>
F;1963-2038/Domain: acyl carrier protein homology <ACP1>
F;2078-2478/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS1>
F;2582-2860/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT1>
F;3710-3891/Domain: short-chain alcohol dehydrogenase homology <SAD2>
F;3995-4066/Domain: acyl carrier protein homology <ACP2>

Query Match 32.7%; Score 6813.5; DB 2; Length 4151;
Best Local Similarity 38.8%; Pred. No. 5.9e-302;
Matches 1668; Conservative 436; Mismatches 1171; Indels 1027; Gaps 91;





```
Db       3247 WGTTEDSAAVVFES--VPLAGDVV----------------AGVYAATSSVLDVLQSW-- 3285

Qy       3387 LERFAGSRLVVVTRGAVVAGPEDGPVDVVGASVWGLVRSAQAEHPDRFVLLDLDTDTGTD 3446

Db       3286 LTRDGAGVLVVMTRGA-VALPGEDVTDLAGAAVWGLVRSAQTEHPGRIVLVDSDAP---- 3340

Qy       3447 LDTGAGAGWGVDGGRVAAVVACGEPQLAVRGERLLAARLK------------------- 3486

Db       3341 ----------LDDSALAAVVTTGEPQVLWRRGEVYTARVHGSRAVGGLLVPPSDRPWRLA 3390

Qy       3487 ----------RLE---------------------------------------------- 3489

Db       3391 MSTAGTFENLRLELIPDADAPLGPGQVRVAVSAIAANFRDVMIALGLYPDPDAVMGVEAC 3450

Qy       3490 -------------SSGD----------------------VPAQRSGDTRARRS------ 3507

Db       3451 GVVIETSLNKGSFAVGDRVMGLFPEGTGTVASTDQRLLVKVPAGWSHTAAATTSVVFATA 3510

Qy       3508 -----DVPAQRS----------GGVP---------------------------------- 3518

Db       3511 HYALVDLAAARSGQRVLIHAGTGGVGMAAVQLARHLGLEVFATASKGKWDTLRAMGFDDD 3570

Qy       3519 ----------------------------ARRSVDVSGREVLP---------------- 3532

Db       3571 HISDSRSLEFEDKFRAATGGRGFDVVLDSLAGEFVDASLRLVAPGGVFLEMGKTDIRDPG 3630

Qy       3533 ------------------------------------------------------------ 3532

Db       3631 VIAQQYPGVRYRAFDLFEPGPDRIAQILAELATLFGDGVLRPLPVTTFDVRCAPAALRYL 3690

Qy       3533 ----------------WLSGGSVLVTGGTGVLGAAVARHLAGVCGVRDLLLVSRRGPDAP 3576

Db       3691 SQARHTGKVVMLMPGSW-AAGTVLITGGTGMAGSAVARHVVARHGVRNLVLVSRRGPDAP 3749

Qy       3577 GAEGLRAELAALGAEVRIVACDVGERREVVRLLEGVPAGCPLTGVVHAAGVLDDATIASL 3636

Db       3750 GAAELVAELAAAGAQVQVVACDAADRAALAKVIADIPVQHPLSGVIHTAGALDDAVVMSL 3809

Qy       3637 TPERLGTVFAAKVDAALLLDELTRGMELSAFVLFSSAAGILGSAGQGNYAAANAALDALA 3696

Db       3810 TPDRVDVVLRSKVDAAWHLHELTRDLDVSAFVMFSSMAGLVGSSGQANYAAANSFLDALA 3869

Qy       3697 YRRRAAGLPGVSLAWGLWEEASGMTGHLAGTDHRRIIRSGLHPMSTPDALALFDAALALD 3756

Db       3870 AHRRAHGLPAISLGWGLWDQASAMTGGLATVDFKRFARDGIVAMSSADALQLFDTAMIVD 3929

Qy       3757 RPVLLPADLRPAP--------PLPPLLQDLLPATRRRTTRTTTTGGADNGAQLHARLAGQ 3808

Db       3930 EPFMLPAHIDFAALKVKFDGGTLPPMFVDLINAPTRRQVDDSLAAAKSKSALLQ-RLEGL 3988

Qy       3809 THEQQHTTLLALVRSHIATVLGHTTPDTIPPDRAFRDLGFDSLTAVELRNRLSRTTGLRL 3868

Db       3989 PEDEQHAVLLDLVRSHIATVLGSASPEAIDPDRAFQELGFDSLTAVEMRNRLKSATGLAL 4048

Qy       3869 PTTLAFDHPNPTTLTHHLHTQL---QPQPDNAVAPVLAELDK---------------LES 3910

Db       4049 SPTLIFDYPNSAALAGYMRRELLGSSPQDTSAVAAGEAELQRIVASIPVKRLRQAGVLDL 4108

Qy       3911 ALSALDKTDSASERVTLRLKSLMLRWNAPQHPTAES--ADDD 3950

Db       4109 LLALANETETSGQDPALA------------PTAEQEIADMD 4137


RESULT 5
T30225
polyketide synthase - Streptomyces hygroscopicus
C;Species: Streptomyces hygroscopicus
C;Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 17-Nov-2000
C;Accession: T30225
R;Aparicio, J.F.; Molnar, I.; Schwecke, T.; Koenig, A.; Haydock, S.F.; Khaw, L.E.; Staun
Gene 169, 9-16, 1996
A;Title: Organization of the biosynthetic gene cluster for rapamycin in Streptomyces hyg
A;Reference number: Z20782; MUID:96186896; PMID:8635756
A;Accession: T30225
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-10223 <APA>
A;Cross-references: EMBL:X86780; NID:g987088; PID:g987099; PIDN:CAA60459.1
C;Genetics:
A;Gene: rapB
C;Superfamily: 3-oxoacyl-[acyl-carrier-protein] synthase I homology; acyl carrier prote
C;Keywords: carrier protein
F;54-449/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS1>
F;1533-1604/Domain: acyl carrier protein homology <ACP1>
F;1647-2042/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS2>
F;2135-2409/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT1>
F;3172-3243/Domain: acyl carrier protein homology <ACP2>
F;3287-3681/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS3>
F;3778-4052/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT2>
F;5160-5231/Domain: acyl carrier protein homology <ACP3>
F;5275-5667/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS4>
F;6776-6847/Domain: acyl carrier protein homology <ACP4>
F;6891-7285/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS5>
F;8411-8482/Domain: acyl carrier protein homology <ACP5>
F;8526-8921/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS6>
F;9012-9285/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT3>
F;10069-10140/Domain: acyl carrier protein homology <ACP>

  Query Match           32.4%;  Score 6756;  DB 2;  Length 10223;
  Best Local Similarity  34.8%;  Pred. No. 8.2e-299;
  Matches 1784;  Conservative  459;  Mismatches 1228;  Indels 1652;  Gaps  118;

Qy         17 VLGVADEADGGVVFVFPGQGPQWPGMGRELLDASDVFRESVRACEAAFAPYVDWSVEQVL 76

Db       2124 VTGTA-VSDPRVVFVFPGQGWQWLGMGSALRDSSIVFAERMAECAPALREFVDWDLFTVL 2182

Qy         77 RDSPDAPGLDRVDVVQPTLFAVMISLAALWRSQGVEPCAVLGHSLGEIAAAHVSGGLSLA 136

Db       2183 -DDPAV--VDRVDVVQPASWRMMVSLAAVWQAAGVRPDAVIGHSQGEIAAACVAGAVSMR 2239

Qy        137 DAARVVTLWSQA-QTTLAGTGALVSVAATPDELLPRIAPWTEDNPARLAVAAVNGPRSTV 195

Db       2240 DAARIVTLRSEAIARGLAGRGAMASVALPAQDVELVDGAW---------IAAHNGPASTV 2290

Qy        196 VSGAREAVADLVADLTAAQVRTRMIPVDVPAHSPLMYAIEERVVSGLLPITPRPSR---- 251

Db       2291 IAGTPEAVDHVLTAHEARGVRVRRITVDYASHTPHVELIRDE----LLDITSDSSSQAPV 2346

Qy        252 IPFHSSVTGGRLDTRELDAAYWYRNMSSTVRFEPAARLLLQQGPKTFVEMSPHPVLTMGL 311

Db       2347 VPWLSTVDGSWVDS-PLDVEYWYRNLREPVGFHPAVGQLQAEGDTVFVEVSASPVLLQAM 2405

Qy        312 QELAPDLGDTTGTADTVIMGTLRRGQGTLDHFLTSLAQLRGHGET--------SATT--- 360

Db       2406 DD------------DVVTVATLRRDDGDATRMLTALAQAYVHGVTVDWPAILGTATTRVL 2453

Qy        361 ---------------------------------------VLSARLTALS---------- 370

Db       2454 DLPTYAFQHQRYWLRSVDRAAADGHPLLGTVVDLPASDGVVLTGRVSLATHTWLADHAVR 2513

Qy        371 ------------------------------------------------------------ 370

Db       2514 GSVLLPGTAFVDLVVRAADEVESDVVDELVIETPLLLPQTGGVQLSVSVGGADESGHRAV 2573

Qy        371 ------------------------------------------------------------ 370

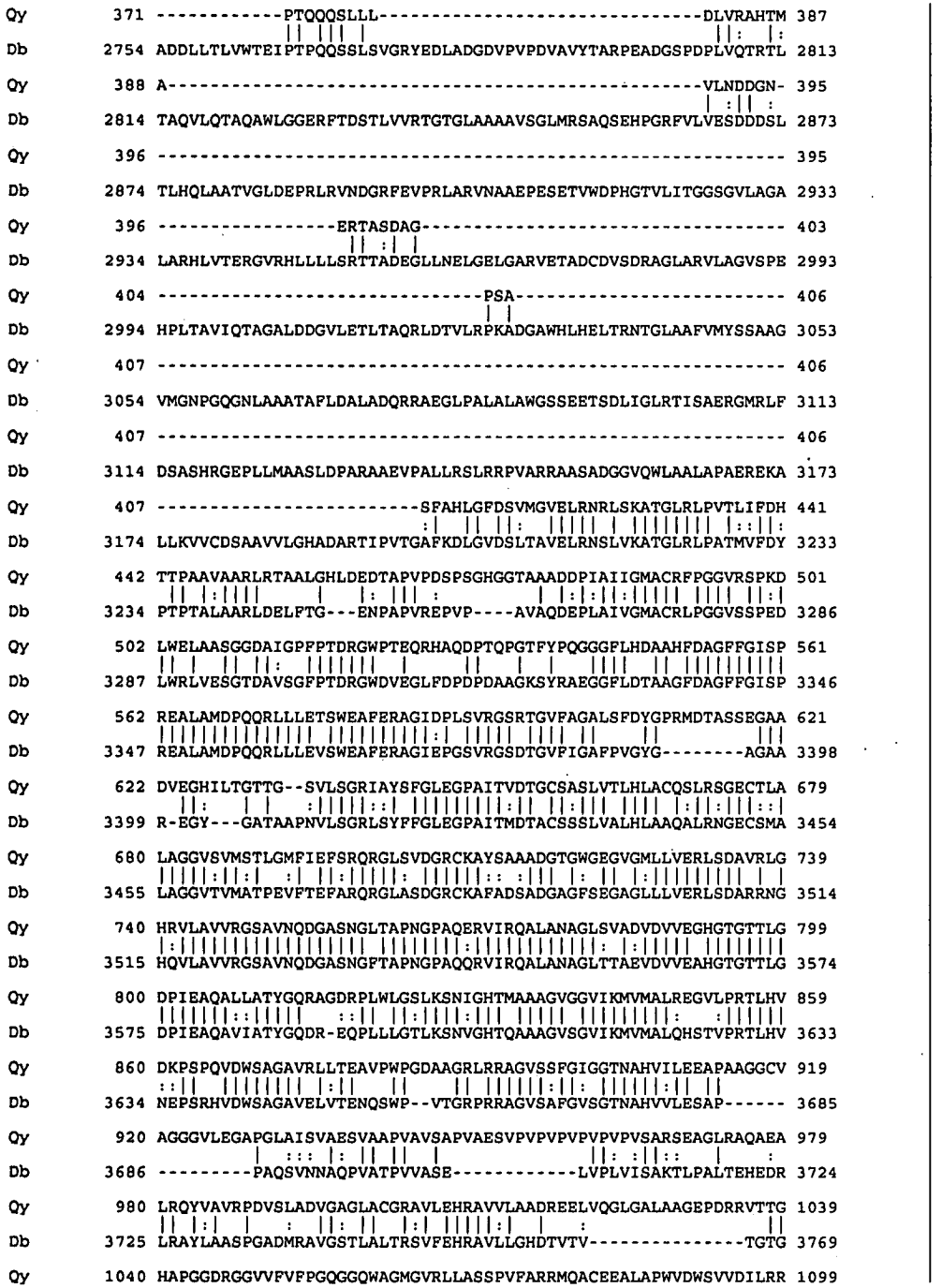
Db       2574 MVFSQADNTDTWTRHVTATVSTSDSTVSLPEFASWPPAQARPVSVADFYDRLAAAGTEYG 2633

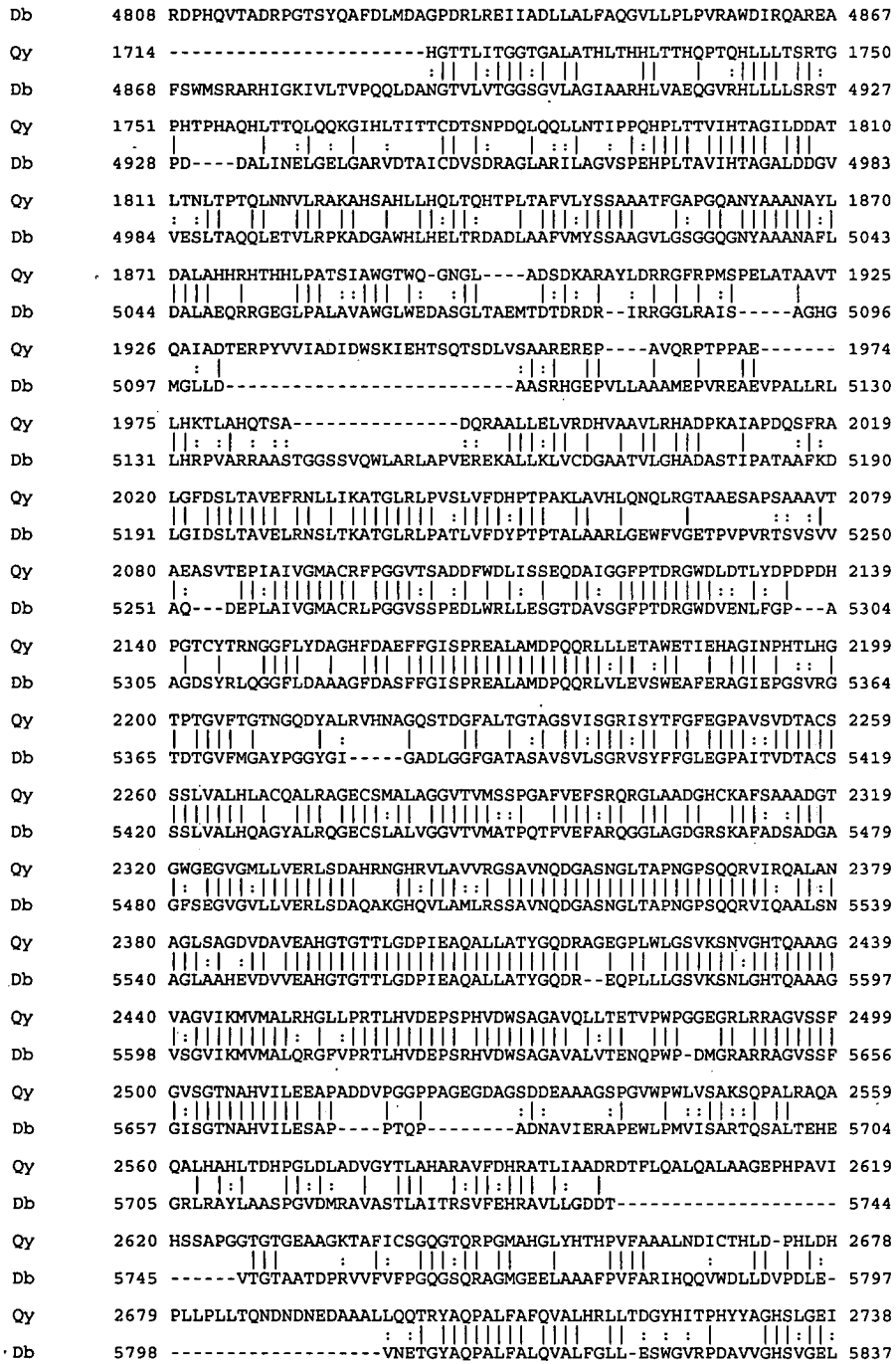
Qy        371 ------------------------------------------------------------ 370

Db       2634 PAFQGLQAAWRDGDTVYAEVVLAEEQAQEAARFAVHPALLDAAMHASVLHTPDTDQQSVR 2693

Qy        371 ------------------------------------------------------------ 370

Db       2694 MPFSWNHVQIRATDTAMLRVAATPTTDGWSVRVADDTGRPVATIGSLVTRPVTADTLGSA 2753
```




Db 4808 RDPHQVTADRPGTSYQAFDLMDAGPDRLREIIADLLALFAQGVLLPLPVRAWDIRQAREA 4867
Qy 1714 ----------HGTTLITGGTGALATHLTHHLTTHQPTQHLLLTSRTG 1750
Db 4868 FSWMSRARHIGKIVLTVPQQLDANGTVLVTGGSGVLAGIAARHLVAEQGVRHLLLLSRST 4927
Qy 1751 PHTPHAQHLTTQLQQKGIHLTITTCDTSNPDQLQQLLNTIPPQHPLTTVIHTAGILDDAT 1810
Db 4928 PD----DALINELGELGARVDTAICDVSDRAGLARILAGVSPEHPLTAVIHTAGALDDGV 4983
Qy 1811 LTNLTPTQLNNVLRAKAHSAHLLHQLTQHTPLTAFVLYSSAAATFGAPGQANYAAANAYL 1870
Db 4984 VESLTAQQLETVLRPKADGAWHLHELTRDADLAAFVMYSSAAGVLGSGGQGNYAAANAFL 5043
Qy 1871 DALAHHRHTHHLPATSIAWGTWQ-GNGL----ADSDKARAYLDRRGFRPMSPELATAAVT 1925
Db 5044 DALAEQRRGEGLPALAVAWGLWEDASGLTAEMTDTDRDR--IRRGGLRAIS-----AGHG 5096
Qy 1926 QAIADTERPYVVIADIDWSKIEHTSQTSDLVSAARAREP----AVQRPTPPAE------- 1974
Db 5097 MGLLD-------------------------AASRHGEPVLLAAAMEPVREAEVPALLRL 5130
Qy 1975 LHKTLAHQTSA---------------DQRAALLELVRDHVAAVLRHADPKAIAPDQSFRA 2019
Db 5131 LHRPVARRAASTGGSSVQWLARLAPVEREKALLKLVCDGAATVLGHADASTIPATAAFKD 5190
Qy 2020 LGFDSLTAVEFRNLLIKATGLRLPVSLVFDHPTPAKLAVHLQNQLRGTAAESAPSAAAVT 2079
Db 5191 LGIDSLTAVELRNSLTKATGLRLPATLVFDYPTPTALAARLGEWFVGETPVPVRTSVSVV 5250
Qy 2080 AEASVTEPIAIVGMACRFPGGVTSADDFWDLISSEQDAIGGFPTDRGWDLDTLYDPDPDH 2139
Db 5251 AQ---DEPLAIVGMACRLPGGVSSPEDLWRLLESGTDAVSGFPTDRGWDVENLFGP---A 5304
Qy 2140 PGTCYTRNGGFLYDAGHFDAEFFGISPREALAMDPQQRLLLETAWETIEHAGINPHTLHG 2199
Db 5305 AGDSYRLQGGFLDAAAGFDASFFGISPREALAMDPQQRLVLEVSWEAFERAGIEPGSVRG 5364
Qy 2200 TPTGVFTGTNGQDYALRVHNAGQSTDGFALTGTAGSVISGRISYTFGFEGPAVSVDTACS 2259
Db 5365 TDTGVFMGAYPGGYGI-----GADLGGFGATASAVSVLSGRVSYFFGLEGPAITVDTACS 5419
Qy 2260 SSLVALHLACQALRAGECSMALAGGVTVMSSPGAFVEFSRQRGLAADGHCKAFSAAADGT 2319
Db 5420 SSLVALHQAGYALRQGECSLALVGGVTVMATPQTFVEFARQGGLAGDGRSKAFADSADGA 5479
Qy 2320 GWGEGVGMLLVERLSDAHRNGHRVLAVVRGSAVNQDGASNGLTAPNGPSQQRVIRQALAN 2379
Db 5480 GFSEGVGVLLVERLSDAQAKGHQVLAMLRSSAVNQDGASNGLTAPNGPSQQRVIQAALSN 5539
Qy 2380 AGLSAGDVDAVEAHGTGTTLGDPIEAQALLATYGQDRAGEGPLWLGSVKSNVGHTQAAAG 2439
Db 5540 AGLAAHEVDVVEAHGTGTTLGDPIEAQALLATYGQDR--EQPLLLGSVKSNLGHTQAAAG 5597
Qy 2440 VAGVIKMVMALRHGLLPRTLHVDEPSPHVDWSAGAVQLLTETVPWPGGEGRLRRAGVSSF 2499
Db 5598 VSGVIKMVMALQRGFVPRTLHVDEPSRHVDWSAGAVALVTENQPWP-DMGRARRAGVSSF 5656
Qy 2500 GVSGTNAHVILEEAPADDVPGGPPAGEGDAGSDDEAAAGSPGVWPWLVSAKSQPALRAQA 2559
Db 5657 GISGTNAHVILESAP----PTQP--------ADNAVIERAPEWLPMVISARTQSALTEHE 5704
Qy 2560 QALHAHLTDHPGLDLADVGYTLAHARAVFDHRATLIAADRDTFLQALQALAAGEPHPAVI 2619
Db 5705 GRLRAYLAASPGVDMRAVASTLAITRSVFEHRAVLLGDDT-------------------- 5744
Qy 2620 HSSAPGGTGTGEAAGKTAFICSGQGTQRPGMAHGLYHTHPVFAAALNDICTHLD-PHLDH 2678
Db 5745 ------VTGTAATDPRVVFVFPGQGSQRAGMGEELAAAFPVFARIHQQVWDLLDVPDLE- 5797
Qy 2679 PLLPLLTQNDNDNEDAAALLQQTRYAQPALFAFQVALHRLLTDGYHITPHYYAGHSLGEI 2738
Db 5798 -------------------VNETGYAQPALFALQVALFGLL-ESWGVRPDAVVGHSVGEL 5837
Qy 2739 TAAHLAGILTLTDATTLITQRATLMQTMPPGTMTTLHTTPHHITHHLTAHENDLAIAAIN 2798
Db 5838 AAGYVSGLWSLEDACTLVSARARLMQALPAGGV--MVAVPVSEDEARAVLGEGVEIAAVN 5895
Qy 2799 TPTSLVISGTPHTVQHITTLCQQQGI-KTKTLPTNHAFHSPHTNPILNQLHQHTQTLTYH 2857
Db 5896 GPSSVVLSGDEAAV-----LQAAEGLGKWTRLATSHAFHSARMEPMLEEFRAVAEGLTYR 5950
Qy 2858 PPHTPLITANTPPDQLLTPHYWTQQARNTVDYATTTQTLHQHGVTTYIELGPDNTLTTLT 2917
Db 5951 TPQVSMAAG----DQLTTTEYWVRQVRDTVRFG---EQVASYEDAVFVELGADRSLARL- 6002
Qy 2918 HHNLPNPPTTTLTLTHPHHHPQ------THLLTNLAKTTTTWHPHHYTHHDNQPHTHTHL 2971
Db 6003 --------VDGVAMLHGDHEAQAAVSALAHLYVN--GVTVDWP----ALLGDAPATRV-L 6047
Qy 2972 DLPTYPFQHHHYWLESTQPGAGNVSAAGLDPTEHPLLGATLELATDGGALLAGRLSLRSH 3031
Db 6048 DLPTYAFQHQRYWLE----GTDRMAAGG-----HPLLGEAVAVPGSDGVLLTGRVSLATH 6098
Qy 3032 PWLADHAVGGTVLLSGATFLELALHAGTYVGCDRVDELTLHAPLVVPVDGGVSVQVGVAA 3091
Db 6099 PWLADHAVRGSVLLPGTGFVELVVRAADEVGCDVIDELVIETPLLLPQTGGVQLSVAVAE 6158
Qy 3092 ADGEGRRLVSVYARGGSACGGGGASGGVWTCHASGVLVEAAAGGVVVDGLAGVWPPRGAV 3151
Db 6159 SDDSGRRAVTVFSRADNV--------DTWTRHAT-VTVSNSDTTLSLPDLS-AWPPAQAQ 6208
Qy 3152 AVDVDGVRDRLAGAGCVLGPVFSGLRAVWRDGGDLLAEVCLPEEAWGDAAGFGLHPALLD 3211
Db 6209 PVNVADFYDQLTVVAYEYGPAFQGLQAAWRDGDTVHAEVALAEEQAQDAARNAVHPALLD 6268
Qy 3212 GVVQPLSVLLPGGTGFGEGAGFGEGVRVPAVWGGVSLHRAGVTGVRVRVSAVGRGGGREA 3271
Db 6269 AALH---------AGILNTPDTEQGVRLPFSWNQVQVHAVGSAMLRVAVTQAADGW---- 6315
Qy 3272 VSVVVGDEAGVPVASVDRLELRPVDMGQLRAVSVSAGRRGSLYAVQWAEVGPVP-----V 3326
Db 6316 -SVRVADDTGRPVATIGSLVTRPVTADTLGSTT------DDLLALTWTEI-PAPQETSLT 6367
Qy 3327 CGQAWAWHEDVGESGGGPVPGVVVLRC-PDAGAGGGGGGGGGGGVGEVVGGVLGVVQGWL 3385
Db 6368 IGR----FEDLADA-DVPVPEVVVFTARPDSSE------NPLAQTRTLTAQVLQKIQAWL 6416
Qy 3386 GLERFAGSRLVVVTRGAVVAGPEDGPVDVVGASVWGLVRSAQAEHPDRFVLLDLDTDTGT 3445
Db 6417 AEERFTDSTLIVRIGTGLAA-----------AAVSGLMRSGQSEHPGRFVLVESDDDT-- 6463
Qy 3446 DLDTGAGAGWGVDGGRVAAVVACGEPQLAVRGERLLAARLKRLESSGDVPAQRSGDTRAR 3505
Db 6464 -----------LAPDQLAATVELDEPRLRISDGRFEVPRL---------------TRTH 6496
Qy 3506 RSDVPAQRSGGVPARRSVDVSGREVLPWLSGGSVLVTGGTGVLGAAVARHLAGVCGVRDL 3565
Db 6497 AAEPESEK-----------------VWDPDGTVLITGGSGVLAGIAARHLVAERGVRHL 6538
Qy 3566 LLVSRRGPDAPGAEGLRAELAALGAEVRIVACDVGERREVVRLLEGVPAGCPLTGVVHAA 3625
Db 6539 LLLSRSAPDG----ALISELGELGAQVATAVCDVSDRPGLVRVLADVSPEHPLTAVIHTA 6594
Qy 3626 GVLDDATIASLTPERLGTVFAAKVDAALLLDELTRGMELSAFVLFSSAAGILGSAGQGNY 3685
Db 6595 GVVDDGVVESLTAQRLDTVLRPKADGAWHLHELTRDADLAAFVMYSSAAGVFGSAGQGNY 6654
Qy 3686 AAANAALDALAYRRRAAGLPGVSLAWGLWEEASGMTGHLAGTDHRRIIRSGLHPMSTPDA 3745
Db 6655 AVANAFLDALAEQRRAEGLPALALAWGLWEGTSGLTANLTDTDHDRIRRSGMRAISAEHG 6714
Qy 3746 LALFDAALALDRPVLLPADLRPA--PPLPPLLQDLLPATRRRTTRTTTTGGADNGAQLHA 3803
Db 6715 MRLFDGASRRRDPVLVAAAMEPVREAEVPAMLRSL----HRPVARRAASGGAVR------ 6764

```
Qy       3804 RLAGQTHEQQHTTLLALVRSHIATVLGHTTPDTIPPDRAFRDLGFDSLTAVELRNRLSRT 3863
Db       6765 CLAALAPEERAKALVKVVCDSAATVLGHADVDSIPVTAAFRDLGVDSLTAVELRNSLTKA 6824

Qy       3864 TGLRLPTTLAFDHPNPTTLTHHL 3886
Db       6825 TGLRLPATLVFDYPTPGALAARL 6847


RESULT 6
T30226
polyketide synthase - Streptomyces hygroscopicus
C;Species: Streptomyces hygroscopicus
C;Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 01-Dec-2000
C;Accession: T30226
R;Aparicio, J.F.; Molnar, I.; Schwecke, T.; Koenig, A.; Haydock, S.F.; Khaw, L.E.; Staun
Gene 169, 9-16, 1996
A;Title: Organization of the biosynthetic gene cluster for rapamycin in Streptomyces hyg
A;Reference number: Z20782; MUID:96186896; PMID:8635756
A;Accession: T30226
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-8563 <APA>
A;Cross-references: EMBL:X86780; NID:g987088; PID:g987100; PIDN:CAA60460.1
C;Genetics:
A;Gene: rapA
C;Superfamily: 3-oxoacyl-[acyl-carrier-protein] synthase I homology; acetate-CoA ligase
C;Keywords: carrier protein; phosphopantetheine; phosphoprotein
F;54-503/Domain: acetate-CoA ligase homology <ACL>
F;1329-1724/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS1>
F;1817-2091/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT1>
F;3199-3270/Domain: acyl carrier protein homology <ACP1>
F;3314-3706/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS2>
F;4787-4858/Domain: acyl carrier protein homology <ACP2>
F;4902-5293/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS3>
F;5386-5659/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT2>
F;6760-6831/Domain: acyl carrier protein homology <ACP3>
F;6875-7269/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS4>
F;7362-7636/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT3>
F;8412-8483/Domain: acyl carrier protein homology <ACP4>
F;4822/Binding site: phosphopantetheine (Ser) (covalent) #status predicted

  Query Match           31.9%;  Score 6642.5;  DB 2;  Length 8563;
  Best Local Similarity  39.2%;  Pred. No. 9.5e-294;
  Matches 1729;  Conservative  427;  Mismatches 1281;  Indels  978;  Gaps  117;

Qy        145 WSQAQTTL--AGTGALVSVAATPDELLPRIAPWTEDNPARLAVAAVNGPRSTVVSGAREA 202
Db       4517 WDPDGTVLITGGSGVLAGIAAR--HLV------TERGVRHLLLLSRSAPDEALIGELGE- 4567

Qy        203 VADLVADLTAAQVRTRMIPVDVPAHSPLMYAIEERVVSGLLPITPRPSRIPFHS------ 256
Db       4568 --------LGARVETAACDVSDPA------ALTQ-VLAGVSPEHPLTAVI--HTAGVVDD 4610

Qy        257 ----SVTGGRLDT---RELDAAYWY-------RNMSSTVRFEPAARLLLQQGPKTF---- 298
Db       4611 GVVESLTVQRLETVLRPKADGA-WNLHELTRDADLAAFVMYSSAAGVLGSAGQANYAAAN 4669

Qy        299 -----------VEMSPHPVLTMGLQE----LAPDLGDTTGTADTVIMGTLR-----RGQG 338
Db       4670 AFLDALAEQRHAEGLPALAVAWGLWEDASGLTAQLTDT--DRDRIRRGGLRAISAEHGMG 4727

Qy        339 TLD----HFLTSL--------------AQLRGHGETSATTVLSAR----LTALSPTQQQS 376
Db       4728 LFDSASRHSEPVLVAAPMEPVRDAEVPALLRSLHRPIARRAAAAGGARWLAALAPAEREK 4787

Qy        377 LLLDLVRAHTMAVLNDDGNERTASDAGPSASFAHLGFDSVMGVELRNRLSKATGLRLPVT 436
Db       4788 ALLKLVCDSAATVL---GHADT-STVSVAAVFRDLGVDSLTAVELRNSLAKATGLRLPAT 4843

Qy        437 LIFDHTTPAAVAARLRTAALGHLDEDTAPVPDSPSGHGGTAAADDPIAIIGMACRFPGGV 496
Db       4844 LVFDYPTPTALAVR-----LGELFTGENPVP--VRGPVSAVAQDEPLAIVGMACRLPGGV 4896

Qy        497 RSPKDLWELAASGGDAIGPFPTDRGWPTEQRHAQDPTQPGTFYPQGGGFLHDAAHFDAGF 556
Db       4897 SSPEDLWRLLESGTDAVSGFPTDRGWDVENLY----DMAGKSHRAEGGFLDAAAGFDAGF 4952

Qy        557 FGISPREALAMDPQQRLLLETSWEAFERAGIDPLSVRGSRTGVFAGALSFDYGPRMDTAS 616
Db       4953 FGISPREALAMDPQQRLVLEVSWEAFERAGIEPGSVRGSDTGVFMGAYPGGYG------- 5005

Qy        617 SEGAADVEGHILTGTTGSVLSGRIAYSFGLEGPAITVDTGCSASLVTLHLACQSLRSGEC 676
Db       5006 --AGADLGGFAATASATSVLSGRVSYFFGLEGPAFTVDTACSSSLVALHQAGYALRQGEC 5063

Qy        677 TLALAGGVSVMSTLGMFIEFSRQRGLSVDGRCKAYSAAADGTGWGEGVGMLLVERLSDAV 736
Db       5064 SLALVGGVTVMATPELFTEFSRQRGLASDGRCKAFADSADGTGWAEGVGVLLVERLSDAQ 5123

Qy        737 RLGHRVLAVVRGSAVNQDGASNGLTAPNGPAQERVIRQALANAGLSVADVDVVEGHGTGT 796
Db       5124 AKGHQVLAVVRSSAVNQDGASNGLTAPNGPSQQRVIQAALSNAGLAAHEVDVVEAHGTGT 5183

Qy        797 TLGDPIEAQALLATYGQRAGDRPLWLGSLKSNIGHTMAAAGVGGVIKMVMALREGVLPRT 856
Db       5184 TLGDPIEAQAVIATYGQDR-ERPLLLGSLKSNIGHAQAASGVSVVIKMVMALQHNTVPRT 5242

Qy        857 LHVDKPSPQVDWSAGAVRLLTEAVPWPGDAAGRLRRAGVSSFGIGGTNAHVILEEAPAAG 916
Db       5243 LHVDEPSRHVDWAAGAVELVRENQPWPG--TDRPRRAGVSSFGVSGTNAHVILESAPPA- 5299

Qy        917 GCVAGGGVLEGAPGLAISVAESVAAPVAVSAPVAESVPVPVPV----PVPVPVSARSEAG 972
Db       5300 ----------------------------QPAEEAQPVETPVVASDVLPLVISAKTQPA 5329

Qy        973 LRAQAEALRQYVAVRPDVSLADVGAGLACGRAVLEHRAVVLAADREELVQGLGALAAGEP 1032
Db       5330 LTEHEDRLRAYLAASPGADTRAVASTLAVTRSVFEHRAVLLGDD--------------- 5373

Qy       1033 DRRVTTGHAPGGDRGGVVFVFPGQGGQWAGMGVRLLASSPVFARRMQACEEALAPWVDW- 1091
Db       5374 ---TVTGTAVSDPR--VVFVFPGQGWQWLGMGSALRDSSVVFAERMAECAAALSEFVDWD 5428

Qy       1092 -SVVDILRRDAGDAVWERADVVQPVLFSVMVSLAALWRSYGIEPDAVLGHSQGEIAAAHV 1150
Db       5429 LTVLD------DPAVVDRVDVVQPASWAVMVSLAAVWQAAGVRPDAVIGHSQGEIAAACV 5482

Qy       1151 CGALSLKDAAKTVALRSRALA-AVRGRGGMASVPLPAQEVEQLIGERWAGRLWVAAVNGP 1209
Db       5483 AGAVSLRDAARIVTLRSQAIARGLAGRGAMASVALPAQDVELVDG------AWIAAHNGP 5536

Qy       1210 RSTAVSGDAEAVDEVLAYCAGTGVRARRIPVDYASHCPHVQPLREELLELLGDISPQPSG 1269
Db       5537 ASTVIAGTPEAVDHVLTAHEARGVRVRRITVDYASHTPHVELIRDELLDITSDSSSQAPL 5596

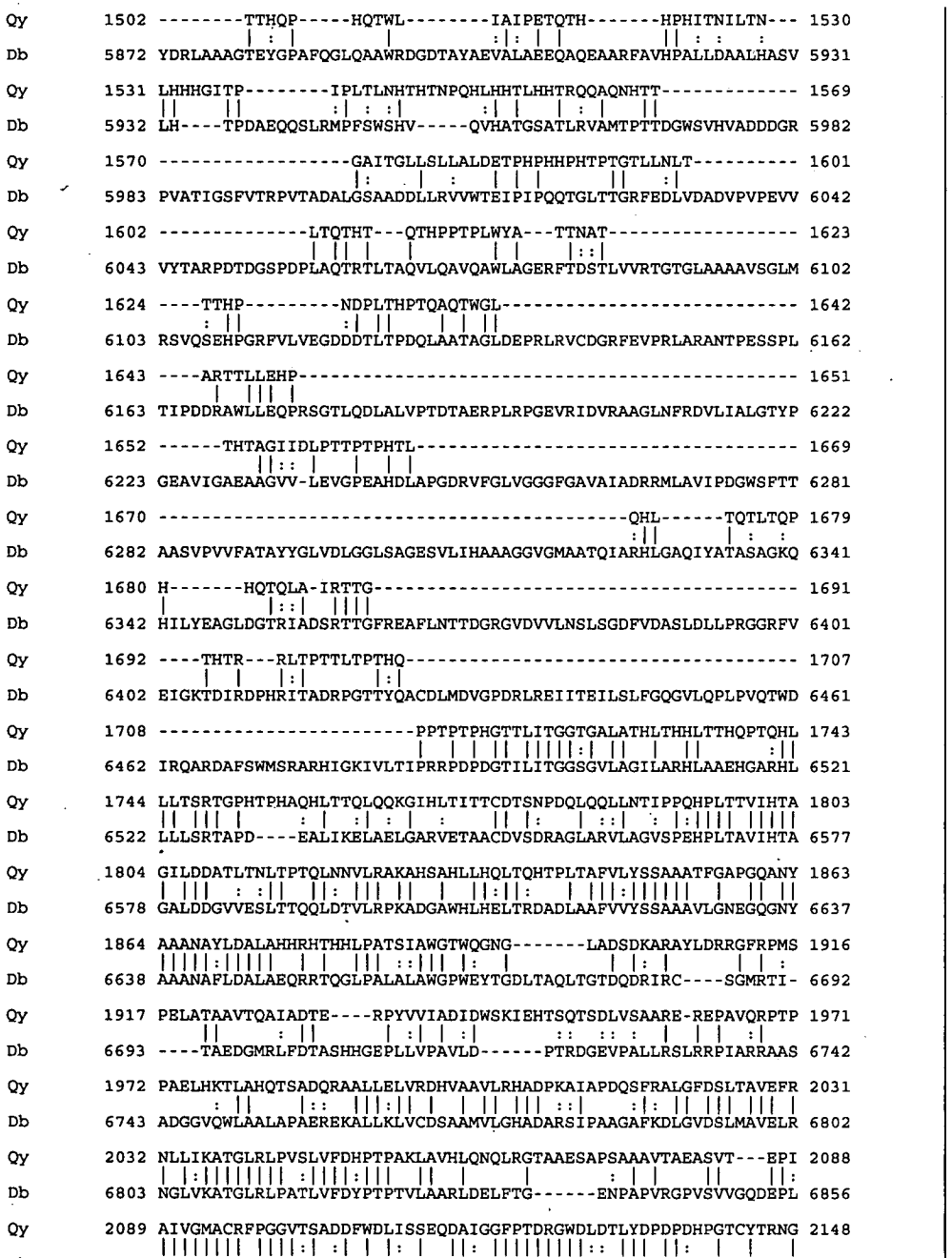
Qy       1270 VPFFSTVEGTWLDTTTLDAAYWYRNLHQPVRFSDAVQALADDGHRVFVEVSPHPTLVPAI 1329
Db       5597 VPWLSTVDGSWVD-SPLDGEYWYRNLREPVGFHPAVGQLQAQGDTVFVEVSASPVLLQAM 5655

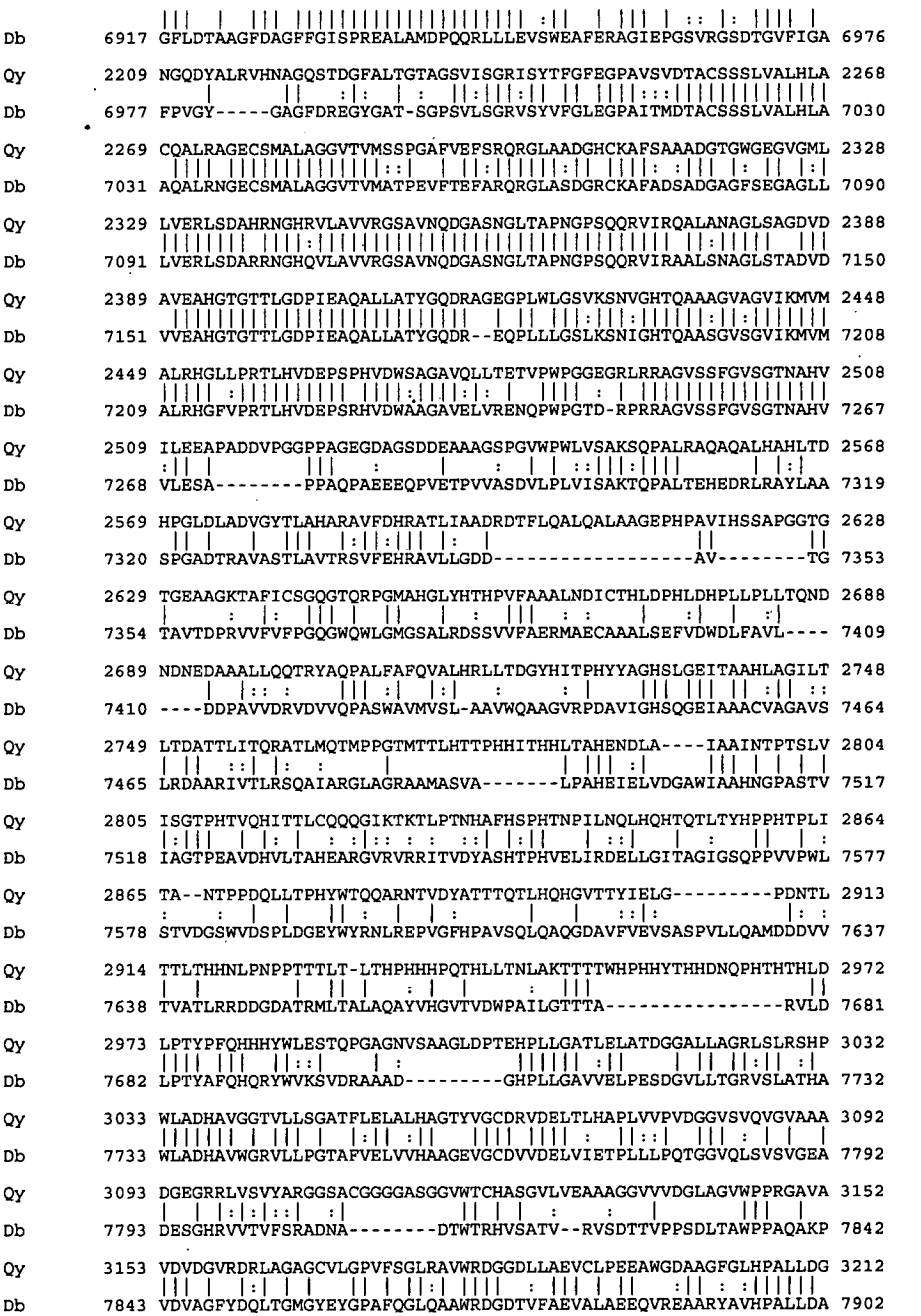
Qy       1330 EDTTEDTAEDVTAIGSLRRGDNDTRRFLTALAHTHTTGI--------GTPTTWHHHYTHH 1381
Db       5656 DD-------DVVTVATLRRDDGDATRMLTALAQAYVHGVTVDWPAILGTTTT-------- 5700

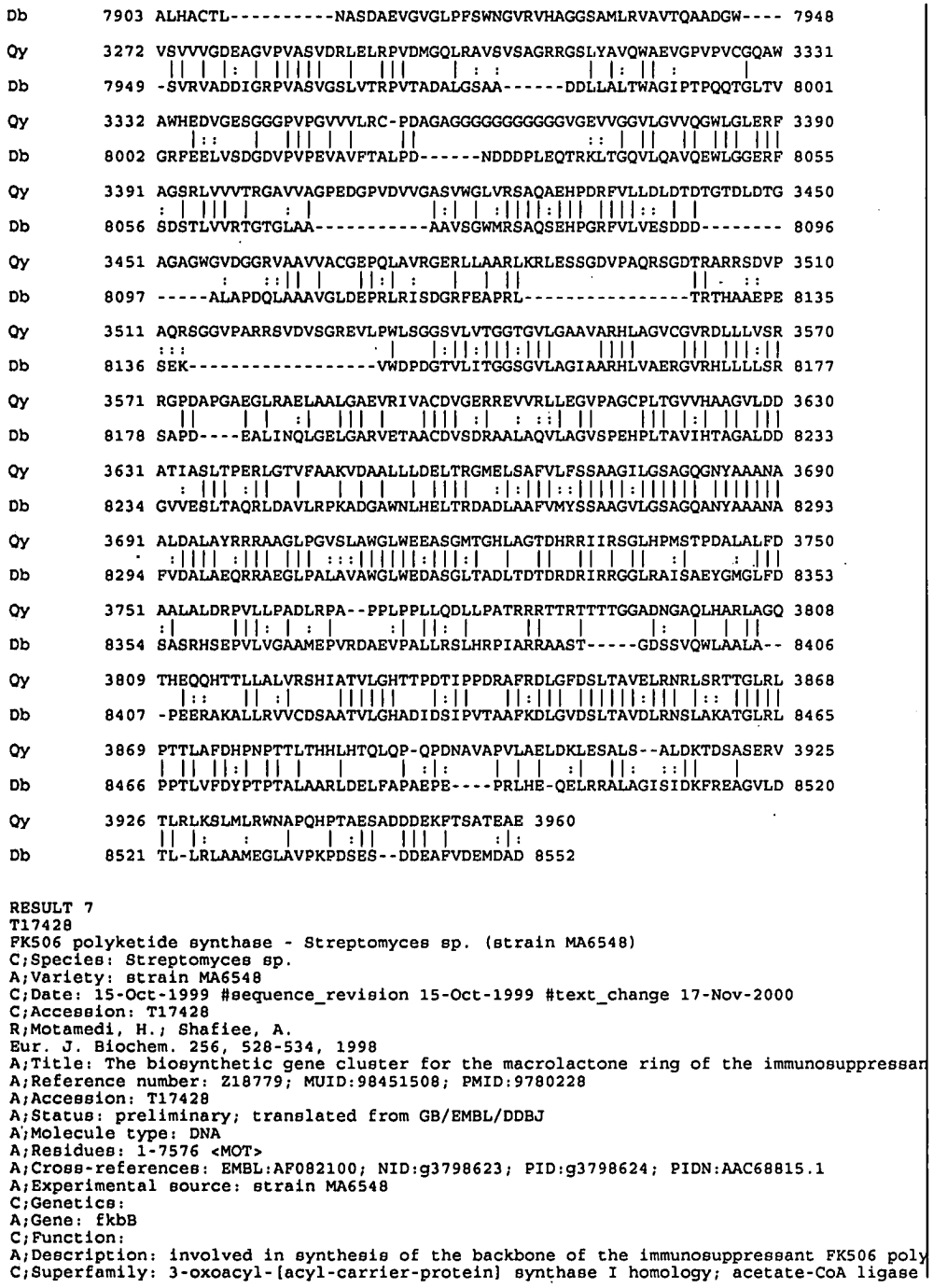
Qy       1382 HTHPHPHTHLDLPTYPFQHQHYWLESSQPGAGSGSG--------AGAGSGAGSGRAGTA- 1432
Db       5701 -------RVLDLPTYAFQHQRYWVEGVDRSAAGGHPLLGVAVELPDSNGVVLTGRVSLAT 5753

Qy       1433 --------------GGTAEVE--SRFWDAVARQDLE--TVATTLAVPPSAGLD---TVV 1470
Db       5754 HTWLADHAVRGSVLLPGTAFVELVVRAADEVECDVIDELVIETPLLLPQTGGVQLSVSVA 5813

Qy       1471 PALSAWHR-----HQHDQARINTW------TYQETWKPLTLP----------------- 1501
Db       5814 EADESGHRAVTVFSRADNA--DTWTRHVSATISASDAPLSLPEFASWPPAQAQPTNVGDL 5871
```



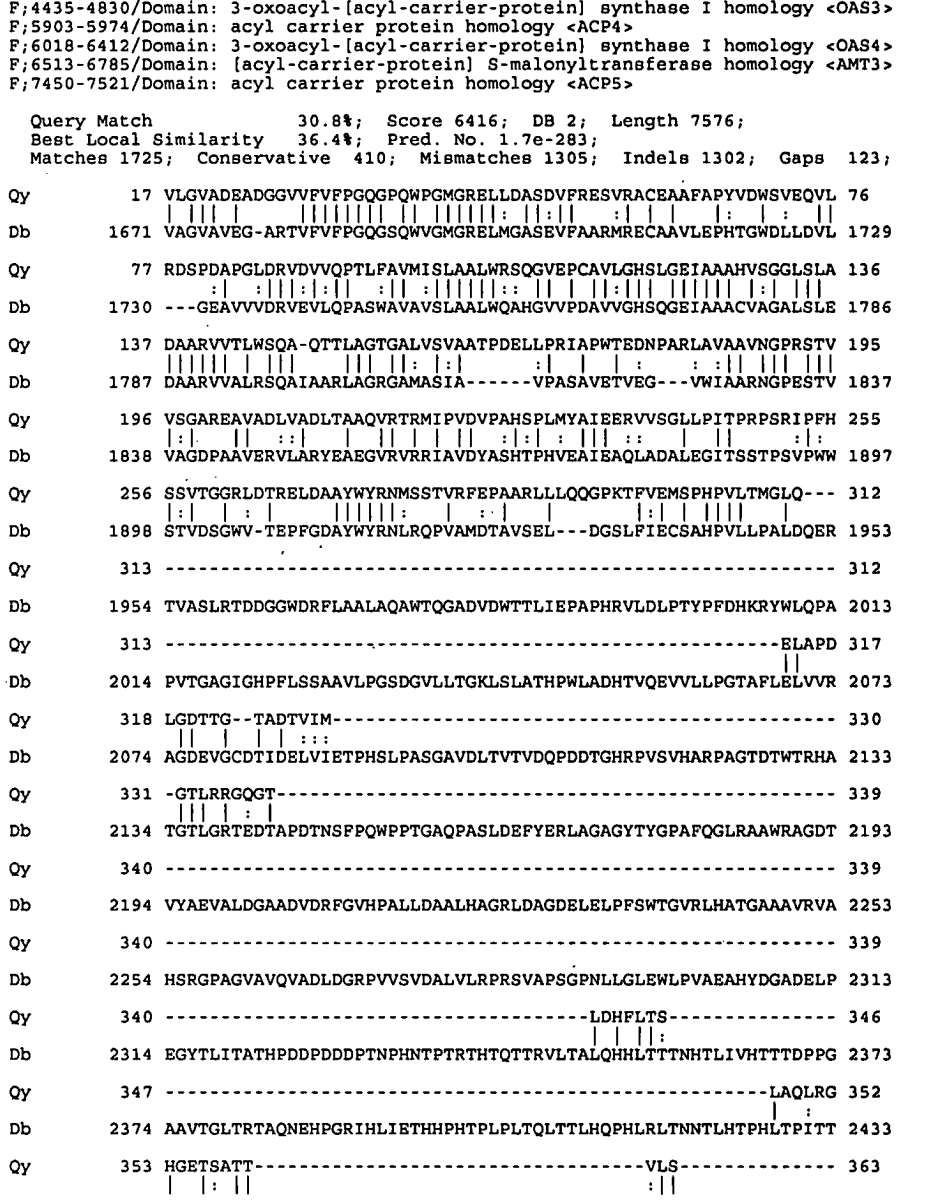


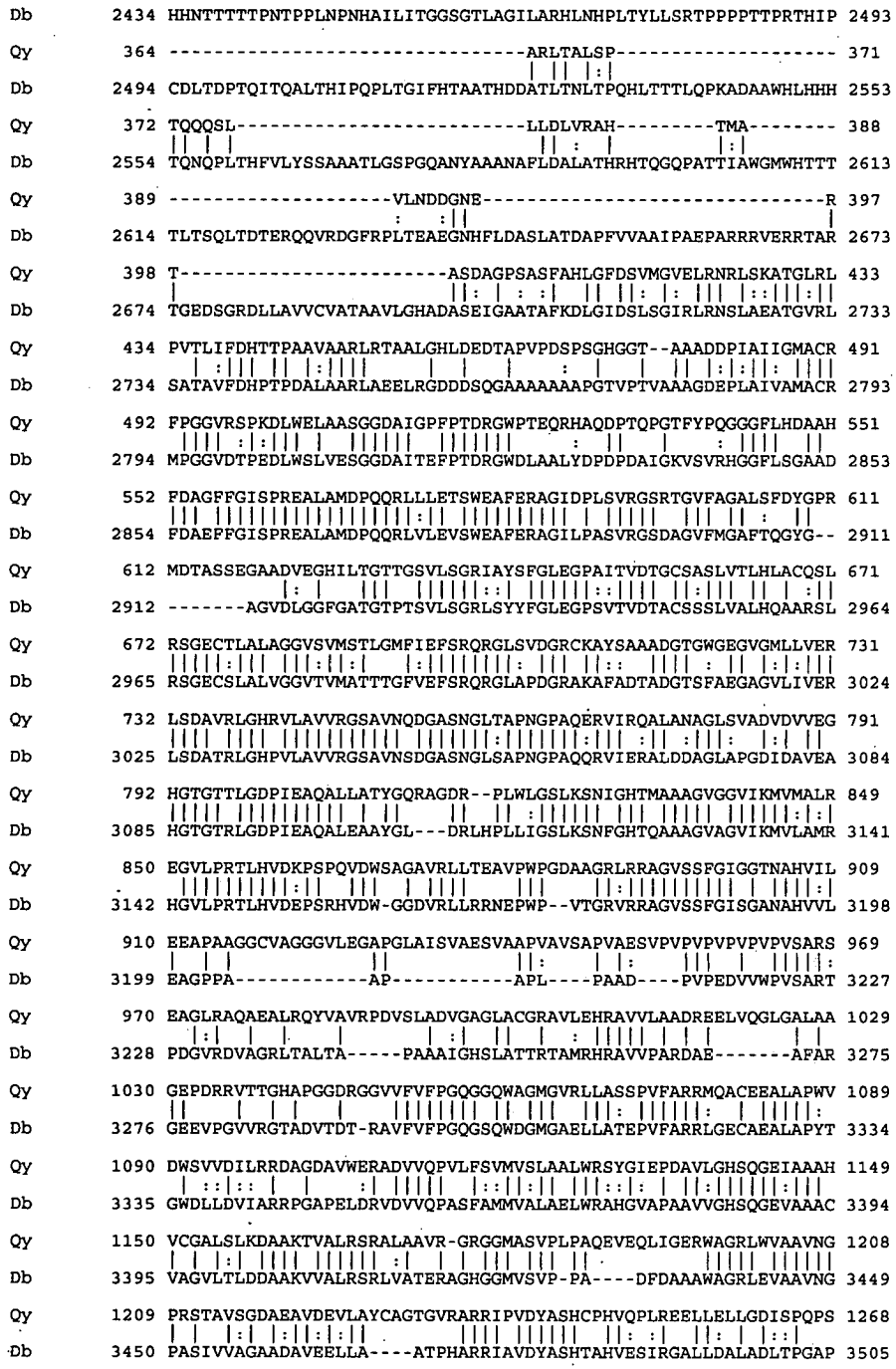


RESULT 7
T17428
FK506 polyketide synthase - Streptomyces sp. (strain MA6548)
C;Species: Streptomyces sp.
A;Variety: strain MA6548
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 17-Nov-2000
C;Accession: T17428
R;Motamedi, H.; Shafiee, A.
Eur. J. Biochem. 256, 528-534, 1998
A;Title: The biosynthetic gene cluster for the macrolactone ring of the immunosuppressan
A;Reference number: Z18779; MUID:98451508; PMID:9780228
A;Accession: T17428
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-7576 <MOT>
A;Cross-references: EMBL:AF082100; NID:g3798623; PID:g3798624; PIDN:AAC68815.1
A;Experimental source: strain MA6548
C;Genetics:
A;Gene: fkbB
C;Function:
A;Description: involved in synthesis of the backbone of the immunosuppressant FK506 poly
C;Superfamily: 3-oxoacyl-[acyl-carrier-protein] synthase I homology; acetate-CoA ligase

C;Keywords: carrier protein
F;54-500/Domain: acetate-CoA ligase homology <ACL>
F;1095-1166/Domain: acyl carrier protein homology <ACP1>
F;1204-1599/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS1>
F;1682-1953/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT1>
F;2680-2751/Domain: acyl carrier protein homology <ACP2>
F;2804-3198/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS2>
F;3295-3569/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT2>
F;4320-4391/Domain: acyl carrier protein homology <ACP3>
F;4435-4830/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS3>
F;5903-5974/Domain: acyl carrier protein homology <ACP4>
F;6018-6412/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS4>
F;6513-6785/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT3>
F;7450-7521/Domain: acyl carrier protein homology <ACP5>

Query Match 30.8%; Score 6416; DB 2; Length 7576;
Best Local Similarity 36.4%; Pred. No. 1.7e-283;
Matches 1725; Conservative 410; Mismatches 1305; Indels 1302; Gaps 123;




Db 2434 HHNTTTTTPNTPPLNPNHAILITGGSGTLAGILARHLNHPLTYLLSRTPPPPTTPRTHIP 2493
Qy 364 ARLTALSP 371
Db 2494 CDLTDPTQITQALTHIPQPLTGIFHTAATHDDATLTNLTPQHLTTTLQPKADAAWHLHHH 2553
Qy 372 TQQQSL LLDLVRAH TMA 388
Db 2554 TQNQPLTHFVLYSSAAATLGSPGQANYAAANAFLDALATHRHTQGQPATTIAWGMWHTTT 2613
Qy 389 VLNDDGNE R 397
Db 2614 TLTSQLTDTERQQVRDGFRPLTEAEGNHFLDASLATDAPFVVAAIPAEPARRRVERRTAR 2673
Qy 398 T ASDAGPSASFAHLGFDSVMGVELRNRLSKATGLRL 433
Db 2674 TGEDSGRDLLAVVCVATAAVLGHADASEIGAATAFKDLGIDSLSGIRLRNSLAEATGVRL 2733
Qy 434 PVTLIFDHTTPAAVAARLRTAALGHLDEDTAPVPDSPSGHGGT AAADDPIAIIGMACR 491
Db 2734 SATAVFDHPTPDALAARLAEELRGDDDSQGAAAAAAAPGTVPTVAAAGDEPLAIVAMACR 2793
Qy 492 FPGGVRSPKDLWELAASGGDAIGPFPTDRGWPTEQRHAQDPTQPGTFYPQGGGFLHDAAH 551
Db 2794 MPGGVDTPEDLWSLVESGGDAITEFPTDRGWDLAALYDPDPDAIGKVSVRHGGFLSGAAD 2853
Qy 552 FDAGFFGISPREALAMDPQQRLLLETSWEAFERAGIDPLSVRGSRTGVFAGALSFDYGPR 611
Db 2854 FDAEFFGISPREALAMDPQQRLVLEVSWEAFERAGILPASVRGSDAGVFMGAFTQGYG 2911
Qy 612 MDTASSEGAADVEGHILTGTTGSVLSGRIAYSFGLEGPAITVDTGCSASLVTLHLACQSL 671
Db 2912 AGVDLGGFGATGTPTSVLSGRLSYYFGLEGPSVTVDTACSSSLVALHQAARSL 2964
Qy 672 RSGECTLALAGGVSVMSTLGMFIEFSRQRGLSVDGRCKAYSAAADGTGWGEGVGMLLVER 731
Db 2965 RSGECSLALVGGVTVMATTTGFVEFSRQRGLAPDGRAKAFADTADGTSFAEGAGVLIVER 3024
Qy 732 LSDAVRLGHRVLAVVRGSAVNQDGASNGLTAPNGPAQERVIRQALANAGLSVADVDVVEG 791
Db 3025 LSDATRLGHPVLAVVRGSAVNSDGASNGLSAPNGPAQQRVIERALDDAGLAPGDIDAVEA 3084
Qy 792 HGTGTTLGDPIEAQALLATYGQRAGDR PLWLGSLKSNIGHTMAAAGVGGVIKMVMALR 849
Db 3085 HGTGTRLGDPIEAQALEAAYGL DRLHPLLIGSLKSNFGHTQAAAGVAGVIKMVLAMR 3141
Qy 850 EGVLPRTLHVDKPSPQVDWSAGAVRLLTEAVPWPGDAAGRLRRAGVSSFGIGGTNAHVIL 909
Db 3142 HGVLPRTLHVDEPSRHVDW GGDVRLLRRNEPWP VTGRVRRAGVSSFGISGANAHVVL 3198
Qy 910 EEAPAAGGCVAGGGVLEGAPGLAISVAESVAAPVAVSAPVAESVPVPVPVPVPVSARS 969
Db 3199 EAGPPA AP APL PAAD PVPEDVVWPVSART 3227
Qy 970 EAGLRAQAEALRQYVAVRPDVSLADVGAGLACGRAVLEHRAVVLAADREELVQGLGALAA 1029
Db 3228 PDGVRDVAGRLTALTA PAAAIGHSLATTRTAMRHRAVVPARDAE AFAR 3275
Qy 1030 GEPDRRVTTGHAPGGDRGGVVFVFPGQGGQWAGMGVRLLASSPVFARRMQACEEALAPWV 1089
Db 3276 GEEVPGVVRGTADVTDT RAVFVFPGQGSQWDGMGAELLATEPVFARRLGECAEALAPYT 3334
Qy 1090 DWSVVDILRRDAGDAVWERADVVQPVLFSVMVSLAALWRSYGIEPDAVLGHSQGEIAAAH 1149
Db 3335 GWDLLDVIARRPGAPELDRVDVVQPASFAMMVALAELWRAHGVAPAAVVGHSQGEVAAAC 3394
Qy 1150 VCGALSLKDAAKTVALRSRALAAVR GRGGMASVPLPAQEVEQLIGERWAGRLWVAAVNG 1208
Db 3395 VAGVLTLDDAAKVVALRSRLVATERAGHGGMVSVP PA DFDAAAWAGRLEVAAVNG 3449
Qy 1209 PRSTAVSGDAEAVDEVLAYCAGTGVRARRIPVDYASHCPHVQPLREELLELLGDISPQPS 1268
Db 3450 PASIVVAGAADAVEELLA ATPHARRIAVDYASHTAHVESIRGALLDALADLTPGAP 3505
Qy 1269 GVPFFSTVEGTWLDTTTLDAAYWYRNLHQPVRFSDAVQALADDGHRVFVEVSPHPTLVPA 1328
Db 3506 EIPFFSTVDEAWLDRPA DAAYWYDNVRCPVRFGAAAARLAELGHRVFVEASPHPVLTTA 3564
Qy 1329 IEDTTEDTAEDVTAI GSLRRGDNDTRRFLTALAHTHTTGIGTPTTWHHHYTHHHTHPHP 1387
Db 3565 LADTL AGHPNTAVTGTLRRGDGGARRFTRSLAELWVRGV PVSW P 3608
Qy 1388 HTHL DLPTYPFQHQHYWLESSQPGAGSGSGAGAGSGAGSGRAGTAGGTAEVESRFWD 1444
Db 3609 FGELRGVPLPTYPFRRDRYWVDAEPAGTSGHPLLGSLVERADGEGAPATALFSVRRQPWL 3668
Qy 1445 AVARQD LETVATTLAVPPSAGLDTVVPALSAWHRHQHDQARINT 1488
Db 3669 ADHEVDGRIIVPGSALVELLAEAGARLGTPTIAELTIVAPVVVDEERDTEIQLTVATEVS 3728
Qy 1489 WTYQET WKPL TLPTTHQPH 1507
Db 3729 GRRAVRLHSRTGAGPWTVGATGALSADTRDPEPVEWPPADAEPVDLTGFYDTLPISYGPA 3788
Qy 1508 QTWLIAIPETQTH HPHITNILTNLHHHGITPIPL TLNHTHTNPQ HLHHTL 1557
Db 3789 FRAMTAMWAGQGRAYASVRLGEQLTDA RYGLHPVLLDAAVHALGTLFADPERRRLAFSW 3847
Qy 1558 HHTRQQAQNHTTGAIT GLLSLLALDE 1583
Db 3848 SGVRIHARAATTQRVLLERADGDTIRILATDEHGAPVLDVDGLTVRAAEPGTEALFEVAW 3907
Qy 1584 TPHP HHPHTPTGTLLNLTLTQTHT QTHPP 1612
Db 3908 VSVPASPVPGWTHLADVPEGDVLEGDAGEGDVGESDPPPVVVLAVEPGDPDSSPGVRARE 3967
Qy 1613 TPLWYATTNATTTHPNDPLTHPTQAQTWGLARTTLLEHPTHTAGII 1658
Db 3968 LGRDLLATVRTWLAAPRWAESRIVVVTRTGD PAQEALGGLVRTAGTEHPGR VGRI 4022
Qy 1659 DLPTTPTPHTLQHLTQTLTQPHHQTQLAIRTTGT HTRRLTPTTLTPTHQPPTPTPH 1714
Db 4023 EADEI TPATVEAGLAAGDESHVRV TDGTVRAARLRRVTAST GASPLSG 4070
Qy 1715 GTTLITGGTGALATHLTHHLTTHQPTQHLLLTSRTG PHTPHAQHLTTQLQQKGIHLTIT 1773
Db 4071 GTVLVTGGTGGLGRLLVDHLLTAHEAAEVVVASRGGRPDGAPADD RVRYV 4120
Qy 1774 TCDTSNPDQLQQLLNTIPPQHPLTTVIHTAGILDDATLTNLTPTQLNNVLRAKAHSAHLL 1833
Db 4121 AADVTDRDELAALVDGVAER LCAVVHMAGIVDDAVVATMRPQQWDAVLRVKADVAWQL 4178
Qy 1834 HQLTQHTPLTAFVLYSSAAATFGAPGQANYAAANAYLDALAHHRHTHHLPATSIAWGTW 1892
Db 4179 HELTRDLELAAFVLYSSISATFGGAGQANYATGNAFLDALARHRRDQGLPAVSLAWGLWD 4238
Qy 1893 QGNGLADSDKARAYLDRRGFRPMSPELATAAVTQAIADTERPYVVIADIDWSKIEH 1948
Db 4239 EADGMGGRLAATDLTR IARGGMIPMTAEQGLELFDAAL HSDRAALVPVRLDLAAVAA 4295
Qy 1949 TSQTSDLVSAARAREPAVQRPTPPAELHKTLAHQTSADQRAALLELVRDHVAAVLRHADP 2008
Db 4296 SDQVPPIL RDLVPNARRAAAP TAARD MLELVRTSAALVLGHRDA 4339
Qy 2009 KAIAPDQSFRALGFDSLTAVEFRNLLIKATGLRLPVSLVFDHPTPAKLAVHLQNQLRG 2066
Db 4340 HAVEPTRAFKEVGFDSLTGVELRNRLADATGLTLPATLVFDHPTAQALAAHL DELAGAR 4398
Qy 2067 TAAESAPSAAAVTAEASVTEPIAIVGMACRFPGGVTSADDFWDLISSEQDAIGGFPTDRG 2126
Db 4399 TATRRRTPAPARRHD EPLAIVGMACRLPGGVASPDDLWRLLESGGDGITAFPADRG 4454
Qy 2127 WDLDTLYDPDPDHPGTCYTRNGGFLYDAGHFDAEFFGISPREALAMDPQQRLLLETAWET 2186
Db 4455 WDVDALYDPDPEHSGTSTVRHGGFLTGAGDFDAGFFGISPREALAMDPQQRVVLETSWEA 4514

```
Qy       2187 IEHAGINPHTLHGTPTGVFTGTNGQDYALRVHNAGQSTDGFALTGTAGSVISGRISYTFG 2246
Db       4515 LEHAGIDPHTLRGTETGVFMG--GYFYG---YGSGADRGDFGATSTQTSVLSGRLSYFYC 4569

Qy       2247 FEGPAVSVDTACSSSLVALHLACQALRAGECSMALAGGVTVMSSPGAFVEFSRQRGLAAD 2306
Db       4570 LEGPAVTVDTACSSSLVALHQAGQSLRTGECSLALVGGVTVMASPSGFVDFSQQRGLAPD 4629

Qy       2307 GHCKAFSAAADGTGWGEGVGMLLVERLSDAHRNGHRVLAVVRGSAVNQDGASNGLTAPNG 2366
Db       4630 GRCKAFAEAADGTAFAEGSGVLVVERLSDAERHGHRVLAVVRGSAVNQDGASNGLSAPNG 4689

Qy       2367 PSQQRVIRQALANAGLSAGDVDAVEAHGTGTTLGDPIEAQALLATYGQDRAGEGPLWLGS 2426
Db       4690 PSQERVIRQALANAGLQPSDVDAIEAHGTGTRLGDPIEATALLATYGQDRA--TPLLLGS 4747

Qy       2427 VKSNVGHTQAAAGVAGVIKMVMALRHGLLPRTLHVDEPSPHVDWSAGAVQLLTETVPWPG 2486
Db       4748 LKSNIGHTQAAAGVAGIIKMVLAMRHGTLPPTLHVDTPSSHVDWSAGTVELLTDARTWP- 4806

Qy       2487 GEGRLRRAGVSSFGVSGTNAHVILEEAPADDVPGGPPAGEGDAGSDDEAAAGSPGVWPWL 2546
Db       4807 ETGRPHRAAVSSFGVSGTNAHVIIESHPR-------PTPAPDTGSSTHPV-------PLL 4852

Qy       2547 VSAKSQPALRAQAQALHAHLTDHPGLDLADVGYTLAHARAVFDHRATLIAADRDTFLQAL 2606
Db       4853 ISARTPEALDEHTTRVRAFL-DAEGEDERAVASALL-TRTAFTHRAALIGTD-------- 4902

Qy       2607 QALAAGEPHPAVIHSSAPGGTGTGEAAGKTAFICSGQGTQRPGMAHGLYHTHPVFAAALN 2666
Db       4903 -----------LI-------TGTAEPDRRLVWLFSGQGSQRPGMGDELAAAYDVFARTRR 4944

Qy       2667 DICTHLDPHLDHPLLPLLTQNDNDNEDAAALLQQTRYAQPALFAFQVALHRLLTDGYHIT 2726
Db       4945 DV-------LDALQVP-----------AGLDVHDTGYAQPAVFALQVALSAQL-DAWGVR 4985

Qy       2727 PHYYAGHSLGEITAAHLAGILTLTDATTLITQRATLMQTMPPGTMTTLHTTPHHITHHLT 2786
Db       4986 PDVLVGHSIGELAAAYVAGVWSLDDATELVSARARLMQALPPGGAMAA----------VS 5035

Qy       2787 AHEND--------LAIAAINTPTSLVISGTPHTVQHITTLCQQQGIKTKTLPTNHAFHSP 2838
Db       5036 ASERDALPLLCEGVEIAAVNGPASIVLSGDEDAVLDVAA---RLGRFTR-LRTSHAFHSA 5091

Qy       2839 HTNPILNQLHQHTQTLTYHPPHTPLITANTPPDQLLTPHYWTQQARNTVDYATTTQTLHQ 2898
Db       5092 RMEPMLDEFRDVAERLTYHEPKLPMAAG----ADCATPEYWVRQVRDTVRFA---EQVAA 5144

Qy       2899 HGVTTYIELGPDNTLTTLTHHNLPNPPTTTLTLTHPHHHPQTHLLTNLAKTTT-----TW 2953
Db       5145 YDGAALLEIGPDRNLARL---------VDGIPVLHGEDEARS-AMTALARLHTGGVAVDW 5194

Qy       2954 HPHHYTHHDNQPHTHTHLDLPTYPFQHHHYWLESTQPGAGNVSAAGLDPTEHPLLGATLE 3013
Db       5195 PEVIGAAPTDLPH------LPTYPFERTRYWLGSR--AAGDAAPAGQLPVAHPVLTAAVM 5246

Qy       3014 LATDGGALLAGRLSLRS-------HPWLADHAVGGTVLLSGATFLELALHAGTYVGCDRV 3066
Db       5247 VPGTGDLVLTGRVEATDPLATRVRHPGAA-----------RAVLLDLRSRRAT----KRA 5291

Qy       3067 DELTLHAP----LVVPVDGGVSVQVGVAAADGEGRRLVSVYARGGSACGGGGASGGVWTC 3122
Db       5292 AAPWRRSPWTPRSALPQSGGLALSVTVAAPGEDGRRAVAVHARPGT---------GEWTE 5342

Qy       3123 HASGVLVEAAAGGVVVDGLAGVWPPRGAVAVDVDGVRDRLAGAGCVLGPVFSGLRAVWRD 3182
Db       5343 HATGILARVARVASAVPETPSPWPPATARPFDVGELADRLARAGHTDGPALPRPRAAWAD 5402

Qy       3183 GGDLLAEVCLPEEAWGDAAGFGLHPALLDGVVQPLSVLLPGGTGFGEGAGFGEGVRVPAV 3242
Db       5403 DDAVHAEVALADEQHADAERYGLHPALLGAAL----------------ALAGEGADLPSA 5446

Qy       3243 WGGVSLHRAGVTGVRVRVSAVGRGGGREAVSVVVGDEAGVPVASVDRLELRPVDMGQLRA 3302
```

```
Db       5447 FDDVRVHATGATTVRVAVTATG---------IHLADETGSPVATVGAVRRRPLTV----- 5492

Qy       3303 VSVSAGRRGSLYAVQWAEVGPVPVCGQAWAWHEDVGESGG---GPVPGVVVLRCPDAGAG 3359
Db       5493 ----EGAVPGLLRPDLAEIAELP---------PTTATTGGLDDPVVPDVVIL------PA 5533

Qy       3360 GGGGGGGGGGVGEVVGGVLGVVQGWLGLERFAGSRLVVVTRGAVVAGPEDGPVDVVGASV 3419
Db       5534 HGTGGGPLGETRDLGARVLAALRSFLTDDRYADAVLAVHT------GPGLAP-----AAA 5582

Qy       3420 WGLVRSAQAEHPDRFVLLDLDTDTGTDLDTGAGAGWGVDGGRVAAVVACGEPQLAVRGER 3479
Db       5583 AGLVRTAQAEHPRRIVVVDAEPDTAAPL-------------LAAAAGLGEPQVVLREGR 5628

Qy       3480 LLAARLKRLESSGDVPAQRSGDTRARRSDVPAQRSGGVPARRSVDVSGREVLPWLSGGSV 3539
Db       5629 AYARRLTPAVPSGDAP--------------------------------ELDP---DGTV 5652

Qy       3540 LVTGGTGVLGAAVARHLAGVCGVRDLLLVSRRG--PDAPGAEGLRAELAALGAEVRIVAC 3597
Db       5653 LITGGSGTLAGIVARHLVGHYGVRRLLMLSRGGTASDVP-----------GAEVTAVAC 5700

Qy       3598 DVGERREVVRLLEGVPAGCPLTGVVHAAGVLDDATIASLTPERLGTVFAAKVDAALLLDE 3657
Db       5701 DIADRDELASALAGIDPAHPLTAVVHTAAVIDDGVLTALTPERLDTVLRPKADGAWHLHE 5760

Qy       3658 LTRGMELSAFVLFSSAAGILGSAGQGNYAA-ANAALDALAYRRRAAGLPGVSLAWGLWEE 3716
Db       5761 LTKDQDLAA--LCSTPRPPECSATRDRATTPRRTLLDELAEQRRAGGLPALSVAWGLWEP 5818

Qy       3717 ASGMTGHLAGTDHRRIIRSGLHPMSTPDALALFDAALALDRPVLLPAD--------LRPA 3768
Db       5819 ESGLT---AGAGD-RMRRDGVTALSTERGLALLDAALRSPEAALVAADPGGLGSSALLRA 5874

Qy       3769 PPLPPLLQDLLPATRRRTTRTTTTGGADNGAQLHARLAGQTHEQQHTTLLALVRSHIATV 3828
Db       5875 PHRGP---------RRR---------AAGASALADRIAGLSAADAEKAALAVVRECAAAV 5916

Qy       3829 LGHTTPDTIPPDRAFRDLGFDSLTAVELRNRLSRTTGLRLPTTLAFDHPNPTTLTHHLHT 3888
Db       5917 LGHEGPARIEATATFKELGVDSLTAVRLRNAFTEATDVRLPATAVFDFPTPQAVAAKLTA 5976

Qy       3889 QL 3890
Db       5977 AL 5978
```

RESULT 8
T43231
probable 6-deoxyerythronolide B synthase, module 1/2 - Saccharopolyspora erythraea
C;Species: Saccharopolyspora erythraea
C;Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 01-Dec-2000
C;Accession: T43231
R;Donadio, S.; Staver, M.J.; McAlpine, J.B.; Swanson, S.J.; Katz, L.
Science 252, 675-679, 1991
A;Title: Modular organization of genes required for complex polyketide biosynthesis.
A;Reference number: Z22354; MUID:91220065; PMID:2024119
A;Accession: T43231
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-3491 <DON>
A;Cross-references: EMBL:M63676; NID:g152691; PID:g152692; PIDN:AAA26493.1
C;Genetics:
A;Gene: eryA
C;Superfamily: 3-oxoacyl-[acyl-carrier-protein] synthase I homology; acyl carrier prote
C;Keywords: carrier protein; phosphopantetheine; phosphoprotein
F;57-336/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT1>
F;416-483/Domain: acyl carrier protein homology <ACP1>
F;526-926/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS1>
F;1031-1317/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT2>
F;1886-1957/Domain: acyl carrier protein homology <ACP2>
F;2001-2400/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS2>

```
F;2508-2791/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT3>
F;3332-3403/Domain: acyl carrier protein homology <ACP3>
F;1921/Binding site: phosphopantetheine (Ser) (covalent) #status predicted

  Query Match             30.5%;  Score 6360;  DB 2;  Length 3491;
  Best Local Similarity   39.3%;  Pred. No. 2.1e-281;
  Matches 1588;  Conservative  435;  Mismatches 1294;  Indels  724;  Gaps   89;

Qy         11 RPAAGEVLGVADEADGGVVFVFPGQGPQWPGMGRELLDASDVFRESVRACEAAFAPYVDW 70
Db         40 RPHASVVRGVA-RPSAPVVFVFPGQGAQWAGMAGELLGESRVFAAAMDACARAFEPVTDW 98

Qy         71 SVEQVLRDSPDAPGLDRVDVVQPTLFAVMISLAALWRSQGVEPCAVLGHSLGEIAAAHVS 130
Db         99 TLAQVL-DSPEQS--RRVEVVQPALFAVQTSLAALWRSFGVTPDAVVGHSIGELAAAHVC 155

Qy        131 GGLSLADAARVVTLWSQAQTTLAGTGALVSVAATPDELLPRIAPWTEDNPARLAVAAVNG 190
Db        156 GAAGAADAARAAALWSREMIPLVGNGDMAAVALSADEIEPRIARWDDD----VVLAGVNG 211

Qy        191 PRSTVVSGAREAVADLVADLTAAQVRTRMIPVDVPAHSPLMYAIEERVVSGLLPITPRPS 250
Db        212 PRSVLLTGSPEPVARRVQELSAEGVRAQVINVSMAAHSAQVDDIAEGMRSALAWFAPGGS 271

Qy        251 RIPFHSSVTGGRLDTRELDAAYWYRNMSSTVRFEPAARLLLQQGPKTFVEMSPHPVLTMG 310
Db        272 EVPFYASLTGGAVDTRELVADYWRRSFRLPVRFDEAIRSALEVGPGTFVEASPHPVLAAA 331

Qy        311 LQELAPDLGDTTGTADTVIMGTLRRGQGTLDHFLTSLAQLRGHGETSATTVLSARLTALS 370
Db        332 LQQTL----DAEGSSAAVV-PTLQRGQGGMRRFLLAAAQAFTGGVAVDWT---AAYDDVG 383

Qy        371 P--------------TQQQSLLLDLVRAHT-----MAVLNDDGNERTASDAGPSASFAHL 411
Db        384 PNPALCRSSRRPRRKTSRPSPASTGTRHRTCCERLLAVVNGETAALAGREADAEATFREL 443

Qy        412 GFDSVMGVELRNRLSKATGLRLPVTLIFDHTTPAAVAARLRTAALGHLDEDTAPVPDSPS 471
Db        444 GLDSVLAAQLRAKVSAAIGREVNIALLYDHPTPRALAEAL---AAG-----TEVAQRETR 495

Qy        472 GHGGTAAADDPIAIIGMACRFPGGVRSPKDLWELAASGGDAIGPFPTDRGWPTEQRHAQD 531
Db        496 ARTNEAAPGEPVAVVAMACRLPGGVSTPEEFWELLSEGRDAVAGLPTDRGWDLDSLFHPD 555

Qy        532 PTQPGTFYPQGGGFLHDAAHFDAGFFGISPREALAMDPQQRLLLETSWEAFERAGIDPLS 591
Db        556 PTRSGTAHQRGGGFLTEATAFDPAFFGMSPREALAVDPQQRLMLELSWEVLERAGIPPTS 615

Qy        592 VRGSRTGVFAGALSFDYGPRMDTASSEGAADVEGHILTGTTGSVLSGRIAYSFGLEGPAI 651
Db        616 LQASPTGVFVGLIPQEYGPRL----AEGGEGVEGYLMTGTTTSVASGRIAYTLGLEGPAI 671

Qy        652 TVDTGCSASLVTLHLACQSLRRGESCTLALAGGVSVMSTLGMFIEFSRQRGLSVDGRCKAY 711
Db        672 SVDTACSSSLVAVHLACQSLRRGESSLAMAGGVTVMPTPGMLVDFSRMNSLAPDGRCKAF 731

Qy        712 SAAADGTGWGEGVGMLLVERLSDAVRLGHRVLAVVRGSAVNQDGASNGLTAPNGPAQERV 771
Db        732 SAGANGFGMAEGAGMLLLERLSDARRNGHPVLAVLRGTAVNSDGASNGLSAPNGRAQVRV 791

Qy        772 IRQALANAGLSVADVDVVEGHGTGTTLGDPIEAQALLATYGQRAGDRPLWLGSLKSNIGH 831
Db        792 IQQALAESGLGPADIDAVEAHGTGTRLGDPIEARALFEAYG-RDREQPLHLGSVKSNLGH 850

Qy        832 TMAAAGVGGVIKMVMALREGVLPRTLHVDKPSPQVDWSAGAVRLLTEAVPWPGDAAGRLR 891
Db        851 TQAAAGVAGVIKMVLAMRAGTLPRTLHASERSKEIDWSSGAISLLDEPEPWP--AGARPR 908

Qy        892 RAGVSSFGIGGTNAHVILEEAPAAGGCVAGGGVLEGAPGLAISVAESVAAPVAVSAPVAE 951
Db        909 RAGVSSFGISGTNAHAIIEEAPQ---------VVEG---------ERVEAGDVVA----- 945

Qy        952 SVPVPVPVPVPVPVSARSEAGLRAQAEALRQYVAVRPDVSLADVGAGLACGRAVLEHRAV 1011

Db        946 ----------PWVLSASSAEGLRAQAARLAAHLREHPGQDPRDIAYSLATGRAALPHRAA 995

Qy       1012 VLAADREELVQGLGALAAGEPDRRVTTGHAPGGDRG--GVVFVFPGQGGQWAGMGVRLLA 1069
Db        996 FAPVDESAALRVLDGLATGNAD-----GAAVGTSRAQQRAVFVFPGQGWQWAGMAVDLLD 1050

Qy       1070 SSPVFARRMQACEEALAPWVDWSVVDILRRDAG----DAV--WERADVVQPVLFSVMVSL 1123
Db        1051 TSPVFAAALRECADALEPHLDFEVIPFLRAEAARREQDAALSTERVDVVQPVMFAVMVSL 1110

Qy       1124 AALWRSYGIEPDAVLGHSQGEIAAAHVCGALSLKDAAKTVALRSRALAAVRGRGGMASVP 1183
Db        1111 ASMWRAHGVEPAAVIGHSQGEIAAACVAGALSLDDAARVVALRSRVIATMPGNKGMASIA 1170

Qy       1184 LPAQEVEQLIGERWAGRLWVAAVNGPRSTAVSGDAEAVDEVLAYCAGTGVRARRIPVDYA 1243
Db        1171 APAGEVRARIGD----RVEIAAVNGPRSVVVAGDSDELDRLVASCTTECIRAKRLAVDYA 1226

Qy       1244 SHCPHVQPLREELLELLG-DISPQPSGVPFFSTVEGTWLDTTTLDAAYWYRNLHQPVRFS 1302
Db        1227 SHSSHVETIRDALHAELGEDFHPLPGFVPFFSTVTGRWTQPDELDAGYWYRNLRRTVRFA 1286

Qy       1303 DAVQALADDGHRVFVEVSPHPTLVPAIEDTTEDTAEDVTAIGSLRRGDNDTRRFLTALAH 1362
Db        1287 DAVRALAEQGYRTFLEVSAHPILTAAIEEIGDGSGADLSAIHSLRRGDGSLADFGEALSR 1346

Qy       1363 THTTGIGTPTTWHHHYTHHHTHPHPHTHLDLPTYPFQHQHYWLESSQPGAGSGSGAGAGS 1422
Db        1347 AFAAGVAVDWESVHLGTGARRVP-------LPTYPFQRERVWLEPK-------------- 1385

Qy       1423 GAGSGRAGTAGGTAEVESRFWDAVARQDLETVATTLAVPPSAGLDTVVPALSAWHRHQHD 1482
Db        1386 ----------------------------------PVARRSTEVDEVSALR----- 1401

Qy       1483 QARINTWTYQETWKPLTLPTTHQPHQTWLIA-----IPETQTHHPHITNILTNLHHHGIT 1537
Db        1402 --------YRIEWRPTGAGEPARLDGTWLVAKYAGTADETSTAAREA------LESAGAR 1447

Qy       1538 PIPLTLNHTHTNPQHLHHTLHHTRQQAQNHTTGAITGLLSLLALDETPHPHHPHTPTGTL 1597
Db        1448 VRELVVD-ARCGRDELAERL---------RSVGEVAGVLSLLAVDEAEPEEAPLALASLA 1497

Qy       1598 LNLTLTQTHTQTHPPTPLWYATTNATTTHPNDPLTHPTQAQTWGLARTTLLEHPTHTAGI 1657
Db        1498 DTLSLVQAMVSAELGCPLWTVTESAVATGPFERVRNAAHGALWGVGRVIALENPAVWGGL 1557

Qy       1658 IDLPTTPTPHTLQHLTQTLTQPHHQTQLAIRTTGTHTRRLTPTTLTPTHQPPTPTPHGTT 1717
Db        1558 VDVPAGSVAELARHLAAVVSGGAGEDQLALRADGVYGRRWVRAAAPATDDEWKPT--GTV 1615

Qy       1718 LITGGTGALATHLTHHLTTHQPTQHLLLTSRTGPHTPHAQHLTTQLQQKGIHLTITTCDT 1777
Db        1616 LVTGGTGGVGGQIARWL-ARRGAPHLLLVSRSGPDADGAGELVAELEALGARTTVAACDV 1674

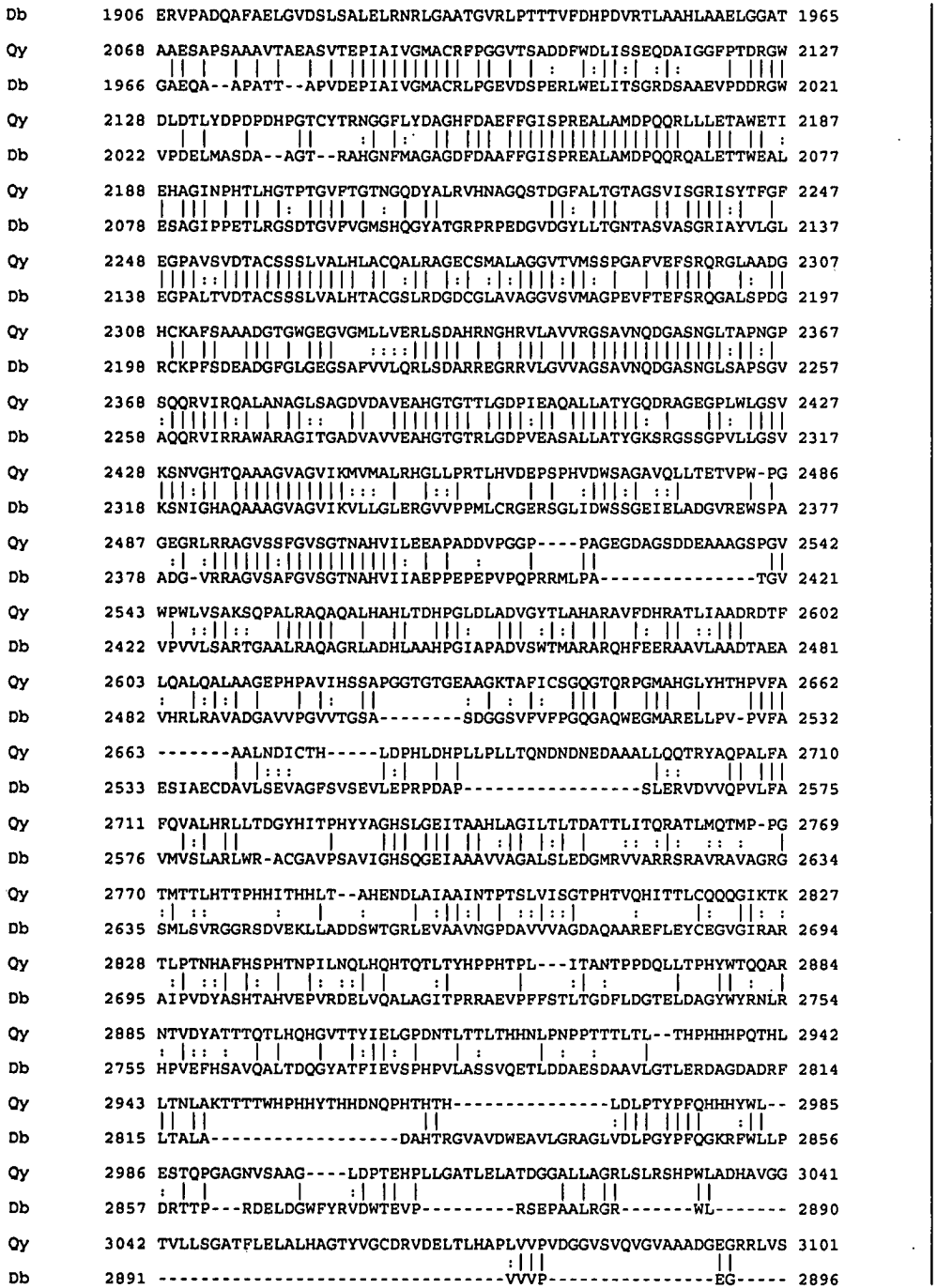
Qy       1778 SNPDQLQQLLNTIPPQHPLTTVIHTAGILDDATLTNLTPTQLNNVLRAKAHSAHLLHQLT 1837
Db        1675 TDRESVRELLGGIGDDVPLSAVFHAAATLDDGTVDTLTGERIERASRAKVLGARNLHELT 1734

Qy       1838 QHTPLTAFVLYSSAAATFGAPGQANYAAANAYLDALAHHRHTHHLPATSIAWGTWQGNGL 1897
Db        1735 RELDLTAFVLFSSFASAFGAPGLGGYAPGNAYLDGLAQQRRSDGLPATAVAWGTWAGSGM 1794

Qy       1898 AD---SDKARAYLDRRGFRPMSPELATAAVTQAIADTERPYVVIADIDWSK--IEHTSQT 1952
Db        1795 AEGAVADRFR----RHGVIEMPPETACRALQNAL-DRAEVCPIVIDVRWDRFLLAYTAQR 1849

Qy       1953 S----DLVSAARREREPAVQRPTPPAELHKTLAHQTSADQRAALLELVRDHVAAVLRHADP 2008
Db        1850 PTRLFDEIDDARRAAP--QAPAEPRV--GALASLPAPEREEALFELVRSHAAAVLGHASA 1905

Qy       2009 KAIAPDQSFRALGFDSLTAVEFRNLLIKATGLRLPVSLVFDHPTPAKLAVHLQNQLRG-T 2067
```



RESULT 9
S23070
erythronolide synthase (EC 2.3.1.94) II - Saccharopolyspora erythraea
N;Alternate names: 6-deoxyerythronolide B synthase II
C;Species: Saccharopolyspora erythraea
C;Date: 07-Apr-1994 #sequence_revision 07-Apr-1994 #text_change 21-Jul-2000
C;Accession: S23070; S22011; S23205
R;Bevitt, D.J.; Cortes, J.; Haydock, S.F.; Leadlay, P.F.
Eur. J. Biochem. 204, 39-49, 1992
A;Title: 6-Deoxyerythronolide-B synthase 2 from Saccharopolyspora erythraea. Cloning of
A;Reference number: S23070; MUID:92155230; PMID:1740151

A;Accession: S23070
A;Molecule type: DNA
A;Residues: 1-3573 <BEV1>
A;Cross-references: EMBL:X62569
A;Experimental source: strain NRRL 2338
R;Bevitt, D.J.
submitted to the EMBL Data Library, September 1991
A;Reference number: S22011
A;Accession: S22011
A;Molecule type: DNA
A;Residues: 1-184,'I',186-301,'S',303-521,523-658,'A',660-993,1001-1212,'H',1214-1392,13
3479,'DH',3480-3572 <BEV2>
A;Cross-references: EMBL:X62569; NID:g46977; PIDN:CAA44448.1; PID:g581651
R;Caffrey, P.; Bevitt, D.J.; Staunton, J.; Leadlay, P.F.
FEBS Lett. 304, 225-228, 1992
A;Title: Identification of DEBS 1, DEBS 2 and DEBS 3, the multienzyme polypeptides of th
A;Reference number: S23103; MUID:92316235; PMID:1618327
A;Accession: S23205
A;Molecule type: protein
A;Residues: 2-12,'XXX' <CAF>
A;Experimental source: strain CA340
C;Genetics:
A;Gene: eryA
A;Start codon: GTG
C;Function:
A;Description: catalyzes the construction of a polyketide chain, which is then cyclised
A;Pathway: erythromycin biosynthesis
C;Superfamily: [acyl-carrier-protein] S-malonyltransferase homology; 3-oxoacyl-[acyl-car
ort-chain alcohol dehydrogenase homology
C;Keywords: acyltransferase; antibiotic biosynthesis; carrier protein; coenzyme A; multi
F;52-453/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS1>
F;561-843/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT1>
F;1140-1308/Domain: short-chain alcohol dehydrogenase homology <SAD2>
F;1404-1475/Domain: acyl carrier protein homology <ACP1>
F;1519-1919/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS2>
F;2023-2305/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT2>
F;2857-3131/Domain: long-chain alcohol dehydrogenase homology <LADH>
F;3149-3327/Domain: short-chain alcohol dehydrogenase homology <SADH>
F;3420-3493/Domain: acyl carrier protein homology <ACP2>
F;1439/Binding site: phosphopantetheine (Ser) (covalent) #status predicted

Query Match 30.2%; Score 6293.5; DB 2; Length 3573;
Best Local Similarity 40.0%; Pred. No. 2.4e-278;
Matches 1526; Conservative 437; Mismatches 1246; Indels 609; Gaps 75;

```
Qy        481 DPIAIIGMACRFPGGVRSPKDLWELAASGGDAIGPFPTDRGWPTEQRHAQDPTQPGTFYP 540
Db         31 DPIAIVSMACRLPGGVNTPQRLWELLREGGETLSGFPTDRGWDLARLHHPDPDNPGTSYV 90

Qy        541 QGGGFLHDAAHFDAGFFGISPREALAMDPQQRLLLETSWEAFERAGIDPLSVRGSRTGVF 600
Db         91 DKGGFLDDAAGFDAEFFGVSPREAAAMDPQQRLLLETSWELVENAGIDPHSLRGTATGVF 150

Qy        601 AGALSFDYGPRMDTASSEGAADVEGHILTGTTGSVLSGRIAYSFGLEGPAITVDTGCSAS 660
Db        151 LGVAKFGYGE--DTAAAE---DVEGYSVTGVAPAVASGRMSYTMGLEGPSISVDTACSSS 205

Qy        661 LVTLHLACQSLRSGECTLALAGGVSVMSTLGMFIEFSRQRGLSVDGRCKAYSAAADGTGW 720
Db        206 LVALHLAVESLRKGESSMAVVGGAAVMATPGVFVDFSRQRALAADGRSKAFGAGADGFGF 265

Qy        721 GEGVGMLLVERLSDAVRLGHRVLAVVRGSAVNQDGASNGLTAPNGPAQERVIRQALANAG 780
Db        266 SEGVTLVLLERLSEARRNGHEVLAVVRGSALNQDGAFNGLSAPSGPAQRRVIRQALESCG 325

Qy        781 LSVADVDVVEGHGTGTTLGDPIEAQALLATYGQ-RAGDRPLWLGSLKSNIGHTMAAAGVG 839
Db        326 LEPGDVDAVEAHGTGTALGDPIEANALLDTYGRDRDADRPLWLGSVKSNIGHTQAAAGVT 385

Qy        840 GVIKMVMALREGVLPRTLHVDKPSPQVDWSAGAVRLLTEAVPWPGDAAGRLRRAGVSSFG 899
Db        386 GLLKVVLALRNGELPATLHVEEPTPHVDWSSGGVALLAGNQPW--RRGERTRRAAVSAFG 443

Qy        900 IGGTNAHVILEEAPAAGGCVAGGGVLEGAPGLAISVAESVAAPVAVSAPVAESVPVPVPV 959
Db        444 ISGTNAHVIVEEAPER----------EHRETTAHDGR---------- 470

Qy        960 PVPVPVSARSEAGLRAQAEALRQYVAVRPDVSLADVGAGLACGRAVLEHR-AVVLAADRE 1018
Db        471 PVPLVVSARSTAALRAQAAQIAELLE-RPDADLAGVGLGLATTRARHEHRAAAVVASTRE 529

Qy       1019 ELVQGLGALAAGEPDRRVTTGHAPGGDRGGVVFVFPGQGGQWAGMGVRLLASSPVFARRM 1078
Db        530 EAVRGLREIAAGAATADAVVEGVTEVDGRNVVFLFPGQGSQWAGMGAELLSSSPVFAGKI 589

Qy       1079 QACEEALAPWVDWSVVDILRRDAGDAVWERADVVQPVLFSVMVSLAALWRSYGIEPDAVL 1138
Db        590 RACDESMAPMQDWKVSDVLRQAPGAPGLDRVDVVQPVLFAVMVSLAELWRSYGVEPAAVV 649

Qy       1139 GHSQGEIAAAHVCGALSLKDAAKTVALRSRALAAVRGRGGMASVPLPAQEVEQLIGERWA 1198
Db        650 GHSQGEIAAEHVAGALTLEDAAKLVVGRSRLMRSLSGEGGMAAVALGEAAVRERL-RPWQ 708

Qy       1199 GRLWVAAVNGPRSTAVSGDAEAVDEVLAYCAGTGVRARRIPVDYASHCPHVQPLREELLE 1258
Db        709 DRLSVAAVNGPRSVVVSGEPGALRAFSEDCAAEGIRVRDIDVDYASHSPQIERVREELLE 768

Qy       1259 LLGDISPQPSGVPFFSTVEGTWLDTTTLDAAYWYRNLHQPVRFSDAVQALADDGHRVFVE 1318
Db        769 TTGDIAPRPARVTFHSTVESRSMDGTELDARYWYRNLRETVRFADAVTRLAESGYDAFIE 828

Qy       1319 VSPHPTLVPAIEDTTE--DTAEDVTAIGSLRRGDNDTRRFLTALAHTHTTGI-------- 1368
Db        829 VSPHPVVVQAVEEAVEEADGAEDAVVVGSLHRDGGDLSAFLRSMATAHVSGVDIRWDVAL 888

Qy       1369 --GTPTTWHHHYTHHHTHPHPHTHLDLPTYPFQHQHYWLESSQPGAGSGSGAGAGSGAGS 1426
Db        889 PGAAP-------------------FALPTYPFQRKRYWLQPAAPAAAS------------ 917

Qy       1427 GRAGTAGGTAEVESRFWDAVARQDLETVATTLAVPPSAGLDTVVPALSAWHRHQHDQARI 1486
Db        918 ----------------DELA---------------------------------------- 921

Qy       1487 NTWTYQETWKPLTLPTTHQPHQTWLIAIPETQTHHPHITNILTN-LHHHGITPIPLTLNH 1545
Db        922 ----YRVSWTPIEKPESGNLDGDWLVVTPLIS---PEWTEMLCEAINANGGRALRCEVD- 973

Qy       1546 THTNPQHLHHTLHHTRQQAQNHTTGAITGLLSLLALDETPHPHHPHTPTGTLLNLTLTQT 1605
Db        974 TSASRTEMAQAVAQAGTGFRQAGTG-FRGVLSLLSSDES--ACRPGVPAGAVGLLTLVQA 1030

Qy       1606 HTQTHPPTPLWYATTNATTTHPNDPLTHPTQAQTWGLARTTLLEHPTHTAGIIDLPTTPT 1665
Db        1031 LGDAGVDAPVWCLTQGAVRTPADDDLARPAQTTAHGFAQVAGLELPGRWGGVVDLPESVD 1090

Qy       1666 PHTLQHLTQTL-----TQPHHQTQLAIRTTGTHTRRLTPTTLTPTHQPPTPTPHGTTLIT 1720
Db        1091 DAALRLLVAVLRGGGRAEDH----LAVRDGRLHGRRVVRASL-PQSGSRSWTPHGTVLVT 1145

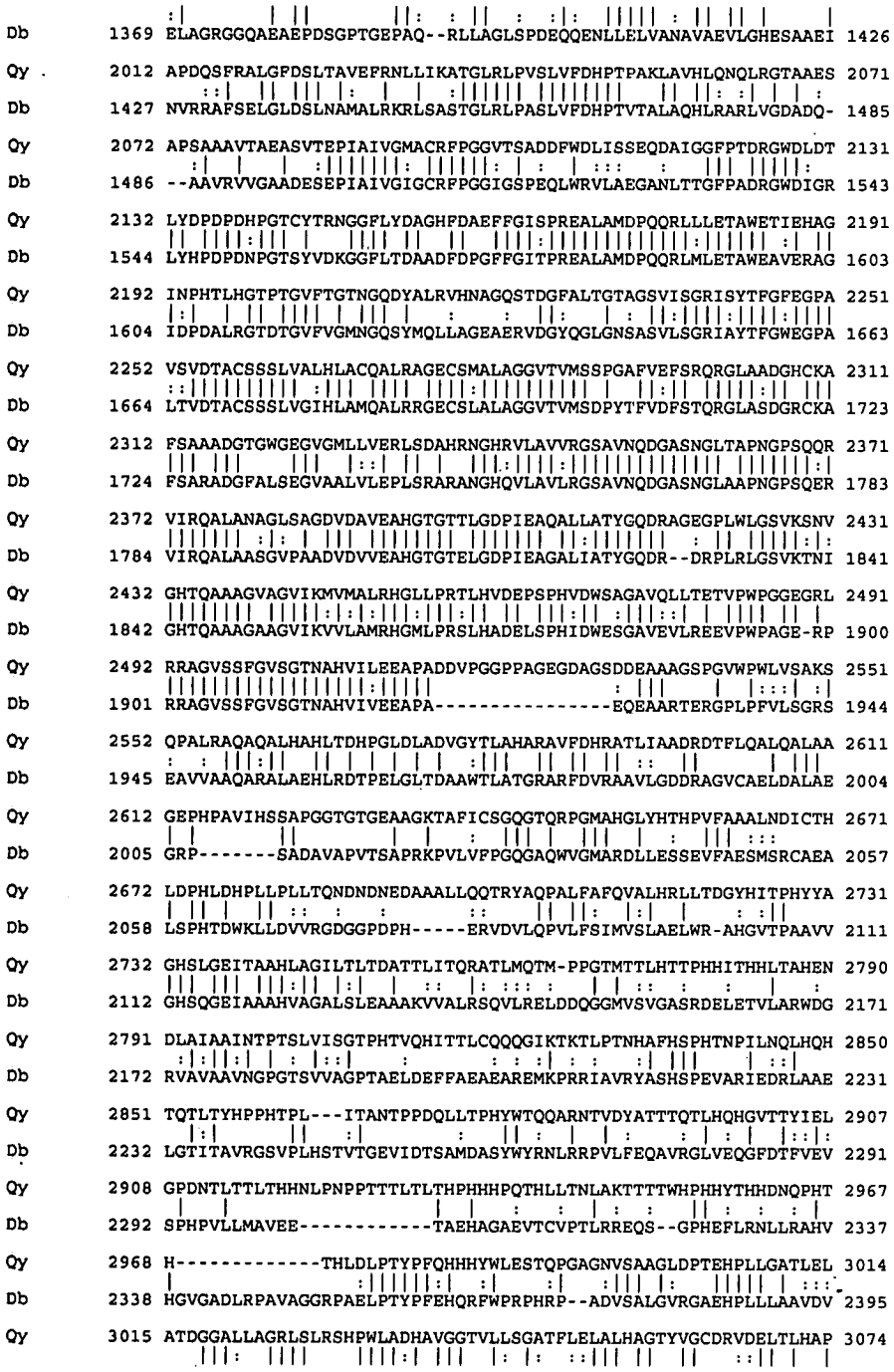
Qy       1721 GGTGALATHLTHHLTTHQPTQHLLLTSRTGPHTPHAQHLTTQLQQKGIHLTITTCDTSNP 1780
Db        1146 GAASPVGDQLVRWLADRGAERLVLAGACPG------DDLLAAVEEAGASAVVCAQDAA-- 1197

Qy       1781 DQLQQLLNTIPPQHPLTTVIHTAGILDDATLTNLTPTQLNNVLRAKAHSAHLLHQLTQHT 1840
Db        1198 -ALREALG----DEPVTALVEAGTLTNFGSISEVAPEEFAETIAAKTALLAVLDEVFGDR 1252

Qy       1841 PLTAFVLYSSAAATFGAPGQANYAAANAYLDALAHHRHTHHLPATSIAWGTWQGNGLADS 1900
Db        1253 AVEREVYCSSVAGIWGGAGMAAYAAGSAYLDALAEHHRARGRSCTSVAWTPWALPGGAVD 1312

Qy       1901 DKARAYLDRRGFRPMSPELATAAVTQAIADTERPYVVIADIDWSKIEH-------TSQTS 1953
Db        1313 D---GYLRERGLRSLSADRAMRTWERVLA-AGPVSVAVADVDWPVLSEGFAATRPTALFA 1368

Qy       1954 DLV--SAARAREREPAVQRPTPPAELHKTLAHQTSADQRAALLELVRDHVAAVLRHADPKAI 2011
```


Db 1369 ELAGRGGQAEAEPDSGPTGEPAQ--RLLAGLSPDEQQENLLELVANAVAEVLGHESAAEI 1426
Qy 2012 APDQSFRALGFDSLTAVEFRNLLIKATGLRLPVSLVFDHPTPAKLAVHLQNQLRGTAAES 2071
Db 1427 NVRRAFSELGLDSLNAMALRKRLSASTGLRLPASLVFDHPTVTALAQHLRARLVGDADQ- 1485
Qy 2072 APSAAAVTAEASVTEPIAIVGMACRFPGGVTSADDFWDLISSEQDAIGGFPTDRGWDLDT 2131
Db 1486 --AAVRVVGAADESEPIAIVGIGCRFPGGIGSPEQLWRVLAEGANLTTGFPADRGWDIGR 1543
Qy 2132 LYDPDPDHPGTCYTRNGGFLYDAGHFDAEFFGISPREALAMDPQQRLLLETAWETIEHAG 2191
Db 1544 LYHPDPDNPGTSYVDKGGFLTDAADFDPGFFGITPREALAMDPQQRLMLETAWEAVERAG 1603
Qy 2192 INPHTLHGTPTGVFTGTNGQDYALRVHNAGQSTDGFALTGTAGSVISGRISYTFGFEGPA 2251
Db 1604 IDPDALRGTDTGVFVGMNGQSYMQLLAGEAERVDGYQGLGNSASVLSGRIAYTFGWEGPA 1663
Qy 2252 VSVDTACSSSLVALHLACQALRAGECSMALAGGVTVMSSPGAFVEFSRQRGLAADGHCKA 2311
Db 1664 LTVDTACSSSLVGIHLAMQALRRGECSLALAGGVTVMSDPYTFVDFSTQRGLASDGRCKA 1723
Qy 2312 FSAAADGTGWGEGVGMLLVERLSDAHRNGHRVLAVVRGSAVNQDGASNGLTAPNGPSQQR 2371
Db 1724 FSARADGFALSEGVAALVLEPLSRARANGHQVLAVLRGSAVNQDGASNGLAAPNGPSQER 1783
Qy 2372 VIRQALANAGLSAGDVDAVEAHGTGTTLGDPIEAQALLATYGQDRAGEGPLWLGSVKSNV 2431
Db 1784 VIRQALAASGVPAADVDVVEAHGTGTELGDPIEAGALIATYGQDR--DRPLRLGSVKTNI 1841
Qy 2432 GHTQAAAGVAGVIKMVMALRHGLLPRTLHVDEPSPHVDWSAGAVQLLTETVPWPGGEGRL 2491
Db 1842 GHTQAAAGAAGVIKVVLAMRHGMLPRSLHADELSPHIDWESGAVEVLREEVPWPAGE-RP 1900
Qy 2492 RRAGVSSFGVSGTNAHVILEEAPADDVPGGPPAGEGDAGSDDEAAAGSPGVWPWLVSAKS 2551
Db 1901 RRAGVSSFGVSGTNAHVIVEEAPA----------EQEAARTERGPLPFVLSGRS 1944
Qy 2552 QPALRAQAQALHAHLTDHPGLDLADVGYTLAHARAVFDHRATLIAADRDTFLQALQALAA 2611
Db 1945 EAVVAAQARALAEHLRDTPELGLTDAAWTLATGRARFDVRAAVLGDDRAGVCAELDALAE 2004
Qy 2612 GEPHPAVIHSSAPGGTGTGEAAGKTAFICSGQGTQRPGMAHGLYHTHPVFAAALNDICTH 2671
Db 2005 GRP-------SADAVAPVTSAPRKPVLVFPGQGAQWVGMARDLLESSEVFAESMSRCAEA 2057
Qy 2672 LDPHLDHPLLPLLTQNDNDNEDAAALLQQTRYAQPALFAFQVALHRLLTDGYHITPHYYA 2731
Db 2058 LSPHTDWKLLDVVRGDGGPDPH-----ERVDVLQPVLFSIMVSLAELWR-AHGVTPAAVV 2111
Qy 2732 GHSLGEITAAHLAGILTLTDATTLITQRATLMQTM-PPGTMTTLHTTPHHITHHLTAHEN 2790
Db 2112 GHSQGEIAAAHVAGALSLEAAAKVVALRSQVLRELDDQGGMVSVGASRDELETVLARWDG 2171
Qy 2791 DLAIAAINTPTSLVISGTPHTVQHITTLCQQQGIKTKTLPTNHAFHSPHTNPILNQLHQH 2850
Db 2172 RVAVAAVNGPGTSVVAGPTAELDEFFAEAEAREMKPRRIAVRYASHSPEVARIEDRLAAE 2231
Qy 2851 TQTLTYHPPHTPL---ITANTPPDQLLTPHYWTQQARNTVDYATTTQTLHQHGVTTYIEL 2907
Db 2232 LGTITAVRGSVPLHSTVTGEVIDTSAMDASYWYRNLRRPVLFEQAVRGLVEQGFDTFVEV 2291
Qy 2908 GPDNTLTTLTHHNLPNPPTTTLTLTHPHHHPQTHLLTNLAKTTTTWHPHHYTHHDNQPHT 2967
Db 2292 SPHPVLLMAVEE-----------TAEHAGAEVTCVPTLRREQS--GPHEFLRNLLRAHV 2337
Qy 2968 H------------THLDLPTYPFQHHHYWLESTQPGAGNVSAAGLDPTEHPLLGATLEL 3014
Db 2338 HGVGADLRPAVAGGRPAELPTYPFEHQRFWPRPHRP--ADVSALGVRGAEHPLLLAAVDV 2395
Qy 3015 ATDGGALLAGRLSLRSHPWLADHAVGGTVLLSGATFLELALHAGTYVGCDRVDELTLHAP 3074
Db 2396 PGHGGAVFTGRLSTDEQPWLAEHVVGGRTLVPGSVLVDLALAAGEDVGLPVLEELVLQRP 2455
Qy 3075 LVVPVDGGVSVQVGVAAADGEGRRLVSVYARGGSACGGGGASGGVWTCHASGVLVEAAAG 3134
Db 2456 LVL-AGAGALLRMSVGAPDESGRRTIDHAAEDVA-----DLADAQWSQHATGTLAQGVAA 2509
Qy 3135 GVVVDGLAGVWPPRGAVAVDVDGVRDRLAGAGCVLGPVFSGLRAVWRDGGDLLAEVCLPE 3194
Db 2510 G---PRDTEQWPPEDAVRIPLDDHYDGLAEQGYEYGPSFQALRAAWRKDDSVYAEVSIA- 2565
Qy 3195 EAWGDAAGFGLHPALLDGVVQPLSVLLPGGTGFGEGAGFGEGVRVPAVWGGVSLHRAGVT 3254
Db 2566 ---ADEEGYAFHPVLLDAVAQTLSLGALGEPG---------GGKLPFAWNTVTLHASGAT 2613
Qy 3255 GVRVRVSAVGRGGGREAVSVVVGDEAGVPVASVDRLELRPVDMGQLRAVSVSAGRRGSLY 3314
Db 2614 SVRV----VATPAGADAMALRVTDPAGHLVATVDSLVVR--STGEKWEQPEPRGGEGELH 2667
Qy 3315 AVQW---AEVGPVPVCGQAWAWHEDVGESGGGPVPGVVVLRC------PDAGAGGGGGGG 3365
Db 2668 ALDWVRLAEPGSTGRVVAADASDLDAVLRSGEPEPDAVLVRYEPEGDDPRAAARHG---- 2723
Qy 3366 GGGGVGEVVGGVLGVVQGWLGLERFAGSRLVVVTRGAVVAGPEDGPVDVVGASVWGLVRS 3425
Db 2724 --------VLWAAALVRRWLEQEELPGATLVIATSGAVTVSDDDSVPEPGAAAMWGVIRC 2775
Qy 3426 AQAEHPDRFVLLDLDT----------DTGTDLDT 3449
Db 2776 AQAESPDRFVLLDTDAEPGMLPAVPDNPQLALRGDDVFVPRLSPLAPSALTLEAGTQRDV 2835
Qy 3450 -GAGA----------GWGVD----------GGRVAA 3464
Db 2836 PGDGAIDSVAFEPAPDVEQPLRRAGEVRVDVRATGVNFRDVLLALGMYPQKADMGTEAAG 2895
Qy 3465 VVACGEPQ----------LAVRGERLLA---------- 3482
Db 2896 VVTAVGPDVDAFPGDRVLGLFQGAFAPIAVTDHRLLARVPDGWSDADAAAVPIAYTTAHY 2955
Qy 3483 ---------- 3482
Db 2956 ALHDLAGLRAGQSVLIHAAAGGVGMAAVALARRAGAEVLATAGPAKHGTLRALGLDDEHI 3015
Qy 3483 --------ARLKR----------LESSGDVPAQ----------RSGDT 3502
Db 3016 ASSRETGFARKFRERTGGRGVDVVLNSLTGELLDESADLLAEDGVFVEMGKTDLRDAGDF 3075
Qy 3503 RAR----------RSDVPAQRSGGVPARRSVDVSGREV- 3530
Db 3076 RGRYAPFDLGEAGDDRLGEILREVVGLLGAGELDRLPVSAWELGSAPAALQHMSRGRHVG 3135
Qy 3531 -------LPWLSGGSVLVTGGTGVLGAAVARHLAGVCGVRDLLLVSRRGPDAPGAEGLRA 3583
Db 3136 MLVLTQPAPVDPDGTVLITGGTGTLGRLLARWLVTEHGVRH-LLVSRRGADAPGSDELRA 3194
Qy 3584 ELAALGAEVRIVACDVGERREVVRLLEGVPAGCPLTGVVHAAGVLDDATIASLTPERLGT 3643
Db 3195 EIEDLGASAEIAACDTADRDALSALLDGLPR--PLTGVVHAAGVLADGLVTSIDEPAVEQ 3252
Qy 3644 VFAAKVDAALLLDELTRGMELSAFVLFSSAAGILGSAGQGNYAAANAALDALAYRRRAAG 3703
Db 3253 VLRAKVDAAWNLHELTANTGLSFFVLFSSAASVLAGPGQGVYAAANESLNALAALRRTRG 3312
Qy 3704 LPGVSLAWGLWEEASGMTGHLAGTDHRRIIRSGLHPMSTPDALALFDAALALDRPVLLP- 3762
Db 3313 LPAKALGWGLWAQASEMTSGLGD----RIARTGVAALPTERALALFDSALRRGGEVVFPL 3368
Qy 3763 ----ADLRPAPPLPPLLQDLLPATRRRTTRTTTTGGAD-NGAQLHARLAGQTHEQQHTTL 3817
Db 3369 SINRSALRRAEFVPEVLRGMVRAKLR------AAGQAEAAGPNVVDRLAGRSESDQVAGL 3422
Qy 3818 LALVRSHIATVLGHTTPDTIPPDRAFRDLGFDSLTAVELRNRLSRTTGLRLPTTLAFDHP 3877
Db 3423 AELVRSHAAAVSGYGSADQLPERKAFKDLGFDSLAAVELRNRLGTATGVRLPSTLVF--P 3480

```
Qy       3878 NPTTLTHHLHTQLQPQPDNAVAPVL---AELDKLESALSALDKTDSASERVTLRLKSLML 3934
Db       3481 TPLAVAEHLRDRL-----FAASPAVDIGDRLDELEKALEALSAEDGHDD-VGQRLESLLR 3534

Qy       3935 RWNA--PQHPTAESADDDEKFTSATEAEIFKFIDNDLG 3970
Db       3535 RWNSRRADAPSTSAISED-----ASDDELFSMLDQRFG 3567
```

RESULT 10
T30228
polyketide synthase - Streptomyces hygroscopicus
C;Species: Streptomyces hygroscopicus
C;Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 17-Nov-2000
C;Accession: T30228
R;Aparicio, J.F.; Molnar, I.; Schwecke, T.; Koenig, A.; Haydock, S.F.; Khaw, L.E.; Staun
Gene 169, 9-16, 1996
A;Title: Organization of the biosynthetic gene cluster for rapamycin in Streptomyces hyg
A;Reference number: Z20782; MUID:96186896; PMID:8635756
A;Accession: T30228
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-6260 <APA>
A;Cross-references: EMBL:X86780; NID:g987088; PID:g987102; PIDN:CAA60462.1
C;Genetics:
A;Gene: rapC
C;Superfamily: 3-oxoacyl-[acyl-carrier-protein] synthase I homology; acyl carrier protei
C;Keywords: carrier protein
F;55-451/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS1>
F;1536-1607/Domain: acyl carrier protein homology <ACP1>
F;1651-2042/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS2>
F;3123-3194/Domain: acyl carrier protein homology <ACP2>
F;3238-3632/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS3>
F;3725-3999/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT1>
F;5114-5185/Domain: acyl carrier protein homology <ACP3>
F;5242-5638/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS4>
F;5759-6044/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT2>
F;6135-6206/Domain: acyl carrier protein homology <ACP4>

 Query Match 29.6%; Score 6164.5; DB 2; Length 6260;
 Best Local Similarity 37.7%; Pred. No. 3.8e-272;
 Matches 1631; Conservative 432; Mismatches 1312; Indels 950; Gaps 116;

```
Qy        145 WSQAQTTL--AGTGALVSVAATPDELLPRIAPWTEDNPARLAVAAVNGPRSTVVSGAREA 202
Db       1266 WDPDGTVLITGGSGVLAGIAAR--HLV------TERGVRHLLLLSRSAPDEALIGELGE- 1316

Qy        203 VADLVADLTAAQVRTRMIPVDVPAHSPLMYAIEERVVSGLLPITPRPSRIPFHS------ 256
Db       1317 --------LGARVETAACDVSDPA------ALTQ-VLAGVSPEHPLTAVI--HTAGVVDD 1359

Qy        257 ----SVTGGRLDT---RELDAAYWY-------RNMSSTVRFEPAARLLLQQGPKTF---- 298
Db       1360 GVVESLTVQRLETVLRPKADGA-WNLHELTRDADLAAFVMYSSAAGVLGSAGQGNYAAAN 1418

Qy        299 -----------VEMSPHPVLTMGLQE----LAPDLGDTTGTADTVIMGTLR-----RGQG 338
Db       1419 AFLDALAEQRHAEGLPALAVAWGLWEDASGLTAQLTDT--DRDRIRRGGLRAISAEHGMG 1476

Qy        339 TLD----HFLTSL--------------AQLRGHGETSATTVLSAR----LTALSPTQQQS 376
Db       1477 LFDSASRHSEPVLVAAPMEPVRDAEVPALLRSLHRPIARRAAAAGGARWLAALAPAEREK 1536

Qy        377 LLLDLVRAHTMAVLNDDGNERTASDAGPSASFAHLGFDSVMGVELRNRLSKATGLRLPVT 436
Db       1537 ALLKLVSDGAATVL---GHADTSTIPATTA-FKDLGINSLTAVELRNSLAKATELRLPAT 1592

Qy        437 LIFDHTTPAAVAARLRTAALGHLDEDTAPVPDSPSGHGGTAAADDPIAIIGMACRFPGGV 496
Db       1593 LVFDYPTPAALAARLDELFTG---ENPVPV----RGPVSAVAQDEPLAIVGMACRLPGGV 1645

Qy        497 RSPKDLWELAASGGDAIGPFPTDRGWPTEQRHAQDPTQPGTFYPQGGGFLHDAAHFDAGF 556
Db       1646 SSPEDLWRLLESGTDAVSGFPTDRGWDVENLY----DMAGKSHRAEGGFLDAAAGFDAGF 1701

Qy        557 FGISPREALAMDPQQRLLLETSWEAFERAGIDPLSVRGSRTGVFAGALSFDYGPRMDTAS 616
Db       1702 FGISPREALAMDPQQRLVLEVSWEAFERAGIEPGSVRGSDTGVFMGAYPGGYGI------ 1755

Qy        617 SEGAADVEGHILTGTTGSVLSGRIAYSFGLEGPAITVDTGCSASLVTLHLACQSLRSGEC 676
Db       1756 ---GADLGGFGATASSVSVLSGRVSYFFGLEGPAFTVDTACSSSLVALHQAGYALRQGEC 1812

Qy        677 TLALAGGVSVMSTLGMFIEFSRQRGLSVDGRCKAYSAAADGTGWGEGVGMLLVERLSDAV 736
Db       1813 SLALVGGVTVMATPQTFVEFSRQGGLASDGRCKAFADAADGTGWAEGVGVLLVERLSDAR 1872

Qy        737 RLGHRVLAVVRGSAVNQDGASNGLTAPNGPAQERVIRQALANAGLSVADVDVVEGHGTGT 796
Db       1873 RNGHQVLAVVRGSAVNQDGASNGLTAPNGPSQQRVIRAALSNAGLSTAEVDVVEAHGTGT 1932

Qy        797 TLGDPIEAQALLATYGQRAGDRPLWLGSLKSNIGHTMAAAGVGGVIKMVMALREGVLPRT 856
Db       1933 TLGDPIEAQALIATYGQDR-DQPVLLGSVKSNLGHTQAAAGVSGVIKMVMALQHGLVPRT 1991

Qy        857 LHVDKPSPQVDWSAGAVRLLTEAVPWPGDAAGRLRRAGVSSFGIGGTNAHVILEEAPAAG 916
Db       1992 LHVDEPSRHVDWSAGAVQLVTENQPWPD--MGRARRAGVSSFGISGTNAHVILESAPPTQ 2049

Qy        917 GCVAGGGVLEGAPGLAISVAESVAAPVAVSAPVAESVPVPVPVPVPVPVSARSEAGLRAQ 976
Db       2050 --PADNAVIERAPEW---------------------------VPLVISARTQSALTEH 2078

Qy        977 AEALRQYVAVRPDVSLADVGAGLACGRAVLEHRAVVLAADREELVQGLGALAAGEPDRRV 1036
Db       2079 EGRLRAYLAASPGVDMRAVASTLAMTRSVFEHRAVLLGDD-------------------T 2119

Qy       1037 TTGHAPGGDRGGVVFVFPGQGGQWAGMGVRLLASSPVFARRMQACEEALAPWVDWSVVDI 1096
Db       2120 VTGTAVSDPR--AVFVFPGQGSQRAGMGEELAAAFPVFARIHQQV---------WDLLDV 2168

Qy       1097 LRRDAGDAVWERADVVQPVLFSVMVSLAALWRSYGIEPDAVLGHSQGEIAAAHVCGALSL 1156
Db       2169 -----PDLEVNETGYAQPALFAMQVALFGLLESWGVRPDAVIGHSVGELAAAYVSGVWSL 2223

Qy       1157 KDAAKTVALRSRALAAVRGRGGMASVPLPAQEVEQLIGERWAGRLWVAAVNGPRSTAVSG 1216
Db       2224 EDACTLVSARARLMQALPAGGVMVAVPVSEDEARAVLGE----GVEIAAVNGPSSVVLSG 2279

Qy       1217 DAEAVDEVLAYCAGTGVRARRIPVDYASHCPHVQPLREELLELLGDISPQPSGVPFFSTV 1276
Db       2280 DEAA---VLQAAEGLG-KWTRLATSHAFHSARMEPMLEEFRAVAEGLTYR---TPQVSMA 2332

Qy       1277 EGTWLDTTTLDAAYWYRNLHQPVRFSDAVQALADDGHRVFVEVSPHPTLVPAIEDTTEDT 1336
Db       2333 VGDQVTT----AEYWVRQVRDTVRFGEQVASYED---AVFVELGADRSLARLVDGV---- 2381

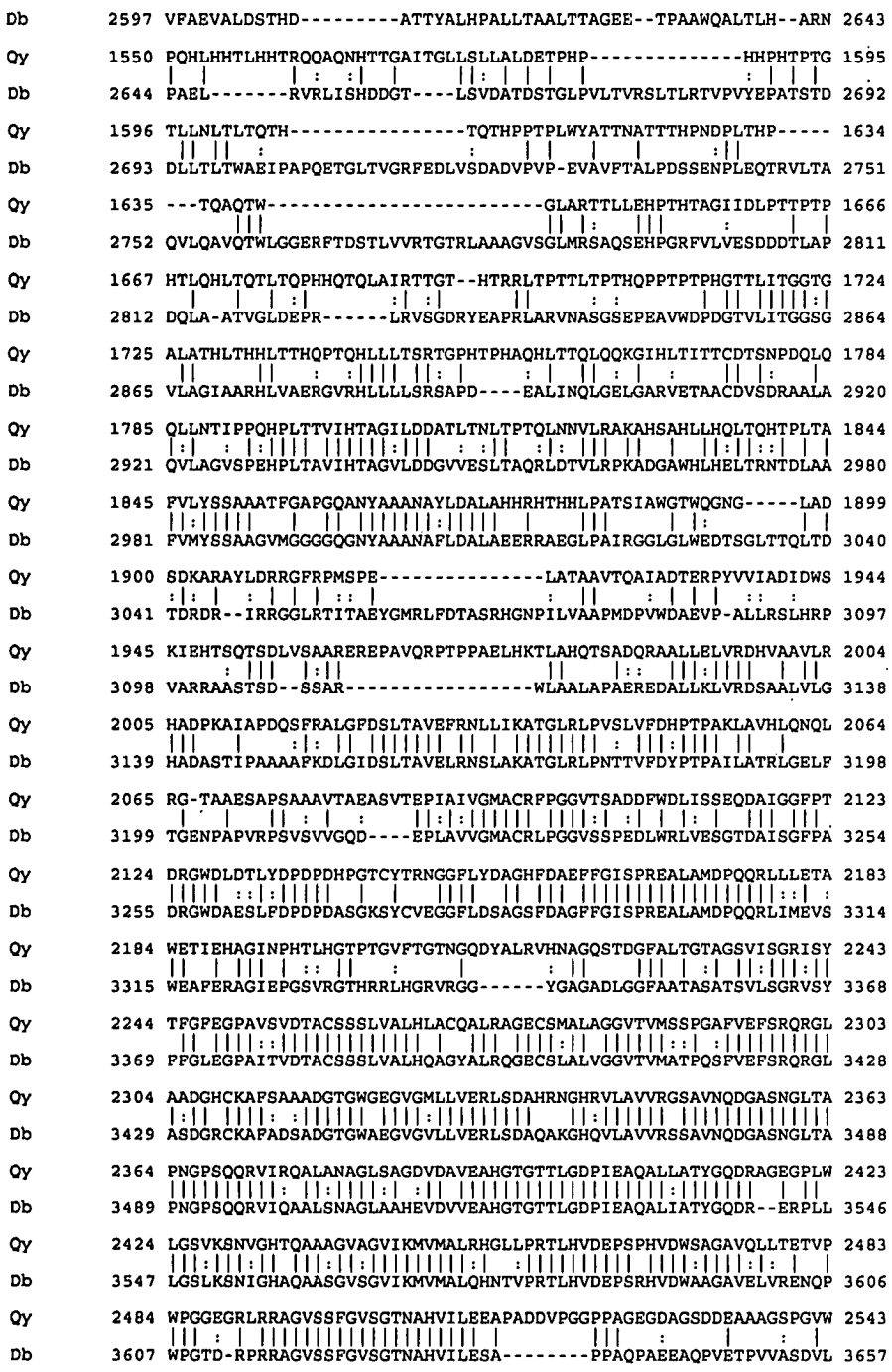
Qy       1337 AEDVTAIGSLRRGDNDTRRFLTALAHTHTTGI---------GTPTTWHHHYTHHHTHPHP 1387
Db       2382 --------AMLHGDHEIQAAIGALAHLYVNGVTVDWPALLGDAPAT-------------- 2419

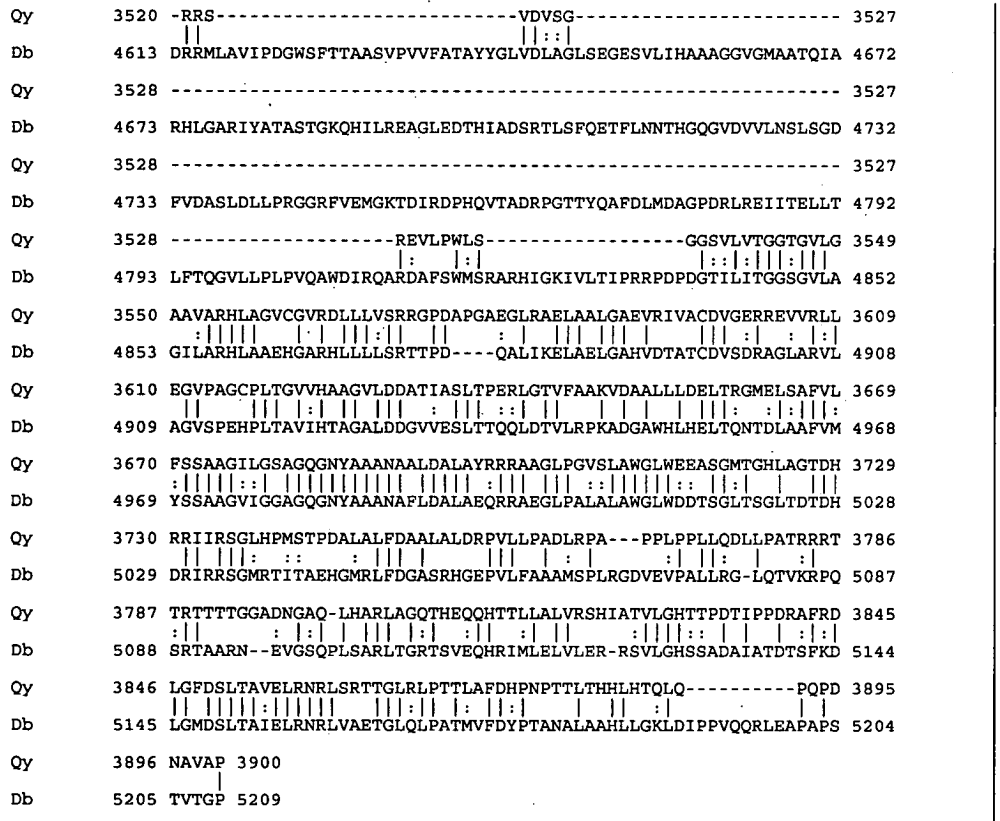
Qy       1388 HTHLDLPTYPFQHQHYWLESSQPGAGSGSGAGAGSGAGSGRAGTAGGTAEVESRFWDAVA 1447
Db       2420 -RVLDLPTYAFQHQRYWLEGTD-RATAGGHPLLGSAVRLAEASGVLFTARV-SRSGDLWL 2476

Qy       1448 RQDLETVAT-----TLAVPPSAGL----DTVVPAL-----------SAW----------- 1476
Db       2477 RDQTVLPATVFVEMALAAADEAGCGLVEDLNVEALLLLPDDGAVQVQTWVSEPDEAGRHR 2536

Qy       1477 ---HRHQHD--------------QARINTWTYQETWKP---------------------- 1497
Db       2537 LSIHARYSDSEPWTRLATATLATRGTVSGWQAGEAWPPTGAVPVETGVPSLRGVWRRGNE 2596

Qy       1498 ----LTLPTTHQPHQTWLIAIPETQTHHPH-ITNILTNLHHHGITPI---PLTLNHTHTN 1549
```

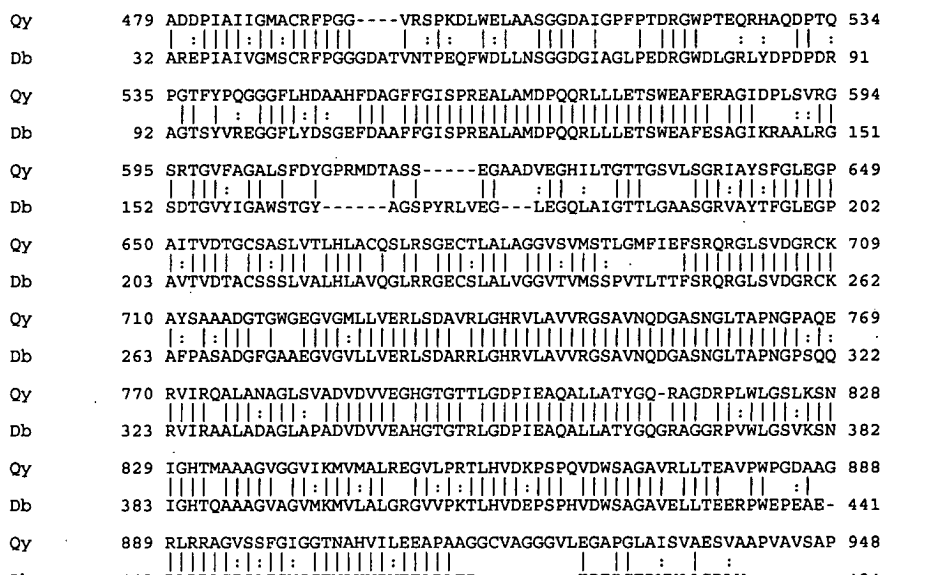

Db 2597 VFAEVALDSTHD---------ATTYALHPALLTAALTTAGEE--TPAAWQALTLH--ARN 2643
Qy 1550 PQHLHHTLHHTRQQAQNHTTGAITGLLSLLALDETPHP----------HHPHTPTG 1595
Db 2644 PAEL-------RVRLISHDDGT----LSVDATDSTGLPVLTVRSLTLRTVPVYEPATSTD 2692
Qy 1596 TLLNLTLTQTH----------TQTHPPTPLWYATTNATTTHPNDPLTHP----- 1634
Db 2693 DLLTLTWAEIPAPQETGLTVGRFEDLVSDADVPVP-EVAVFTALPDSSENPLEQTRVLTA 2751
Qy 1635 ---TQAQTW----------GLARTTLLEHPTHTAGIIDLPTTPTP 1666
Db 2752 QVLQAVQTWLGGERFTDSTLVVRTGTRLAAAGVSGLMRSAQSEHPGRFVLVESDDDTLAP 2811
Qy 1667 HTLQHLTQTLTQPHHQTQLAIRTTGT--HTRRLTPTTLTPTHQPPTPTPHGTTLITGGTG 1724
Db 2812 DQLA-ATVGLDEPR------LRVSGDRYEAPRLARVNASGSEPEAVWDPDGTVLITGGSG 2864
Qy 1725 ALATHLTHHLTTHQPTQHLLLTSRTGPHTPHAQHLTTQLQQKGIHLTITTCDTSNPDQLQ 1784
Db 2865 VLAGIAARHLVAERGVRHLLLLSRSAPD----EALINQLGELGARVETAACDVSDRAALA 2920
Qy 1785 QLLNTIPPQHPLTTVIHTAGILDDATLTNLTPTQLNNVLRAKAHSAHLLHQLTQHTPLTA 1844
Db 2921 QVLAGVSPEHPLTAVIHTAGVLDDGVVESLTAQRLDTVLRPKADGAWHLHELTRNTDLAA 2980
Qy 1845 FVLYSSAAATFGAPGQANYAAANAYLDALAHHRHTHHLPATSIAWGTWQGNG-----LAD 1899
Db 2981 FVMYSSAAGVMGGGGQGNYAAANAFLDALAEERRAEGLPAIRGGLGLWEDTSGLTTQLTD 3040
Qy 1900 SDKARAYLDRRGFRPMSPE----------LATAAVTQAIADTERPYVVIADIDWS 1944
Db 3041 TDRDR--IRRGGLRTITAEYGMRLFDTASRHGNPILVAAPMDPVWDAEVP-ALLRSLHRP 3097
Qy 1945 KIEHTSQTSDLVSAARELEPAVQRPTPPAELHKTLAHQTSADQRAALLELVRDHVAAVLR 2004
Db 3098 VARRAASTSD--SSAR----------WLAALAPAEREDALLKLVRDSAALVLG 3138
Qy 2005 HADPKAIAPDQSFRALGFDSLTAVEFRNLLIKATGLRLPVSLVFDHPTPAKLAVHLQNQL 2064
Db 3139 HADASTIPAAAAFKDLGIDSLTAVELRNSLAKATGLRLPNTTVFDYPTPAILATRLGELF 3198
Qy 2065 RG-TAAESAPSAAAVTAEASVTEPIAIVGMACRFPGGVTSADDFWDLISSEQDAIGGFPT 2123
Db 3199 TGENPAPVRPSVSVVGQD----EPLAVVGMACRLPGGVSSPEDLWRLVESGTDAISGFPA 3254
Qy 2124 DRGWDLDTLYDPDPDHPGTCYTRNGGFLYDAGHFDAEFFGISPREALAMDPQQRLLLETA 2183
Db 3255 DRGWDAESLFDPDPDASGKSYCVEGGFLDSAGSFDAGFFGISPREALAMDPQQRLIMEVS 3314
Qy 2184 WETIEHAGINPHTLHGTPTGVFTGTNGQDYALRVHNAGQSTDGFALTGTAGSVISGRISY 2243
Db 3315 WEAFERAGIEPGSVRGTHRRLHGRVRGG------YGAGADLGGFAATASATSVLSGRVSY 3368
Qy 2244 TFGPEGPAVSVDTACSSSLVALHLACQALRAGECSMALAGGVTVMSSPGAFVEFSRQRGL 2303
Db 3369 PFGLEGPAITVDTACSSSLVALHQAGYALRQGECSLALVGGVTVMATPQSFVEFSRQRGL 3428
Qy 2304 AADGHCKAFSAAADGTGWGEGVGMLLVERLSDAHRNGHRVLAVVRGSAVNQDGASNGLTA 2363
Db 3429 ASDGRCKAFADSADGTGWAEGVGVLLVERLSDAQAKGHQVLAVVRSSAVNQDGASNGLTA 3488
Qy 2364 PNGPSQQRVIRQALANAGLSAGDVDAVEAHGTGTTLGDPIEAQALLATYGQDRAGEGPLW 2423
Db 3489 PNGPSQQRVIQAALSNAGLAAHEVDVVEAHGTGTTLGDPIEAQALIATYGQDR--ERPLL 3546
Qy 2424 LGSVKSNVGHTQAAAGVAGVIKMVMALRHGLLPRTLHVDEPSPHVDWSAGAVQLLTETVP 2483
Db 3547 LGSLKSNIGHAQAASGVSGVIKMVMALQHNTVPRTLHVDEPSRHVDWAAGAVELVRENQP 3606
Qy 2484 WPGGEGRLRRAGVSSFGVSGTNAHVILEEAPADDVPGGPPAGEGDAGSDDEAAAGSPGVW 2543
Db 3607 WPGTD-RPRRAGVSSFGVSGTNAHVILESA--------PPAQPAEEAQPVETPVVASDVL 3657
Qy 2544 PWLVSAKSQPALRAQAQALHAHLTDHPGLDLADVGYTLAHARAVFDHRATLIAADRDTFL 2603
Db 3658 PLVISAKTQPALTEHEDRLRAYLAASPGADIRAVASTLAVTRSVFEHRAVLLGDDT---- 3713
Qy 2604 QALQALAAGEPHPAVIHSSAPGGTGTGEAAGKTAFICSGQGTQRPGMAHGLYHTHPVFAA 2663
Db 3714 ----------VTGTAVTDPRIVFVFPGQGWQWLGMGSALRDSSVVFAE 3751
Qy 2664 ALNDICTHLDPHLDHPLLPLLTQNDNDNEDAAALLQQTRYAQPALFAFQVALHRLLTDGY 2723
Db 3752 RMAECAAALREFVDWDLFTVL--------DDPAVVDRVDVVQPASWAMMVSL-AAVWQAA 3802
Qy 2724 HITPHYYAGHSLGEITAAHLAGILTLTDATTLITQR--ATLMQTMPPGTMTTLHTTPHHI 2781
Db 3803 GVRPDAVIGHSQGEIAAACVAGAVSLRDAARIVTLRSQAIARGLAGRGAMASV------- 3855
Qy 2782 THHLTAHENDLA----IAAINTPTSLVISGTPHTVQHITTLCQQQGIKTKTLPTNHAFHS 2837
Db 3856 --ALPAQDVELVDGAWIAAHNGPASTVIAGTPEAVDHVLTAHEAQGVRVRRITVDYASHT 3913
Qy 2838 PHTNPILNQLHQHTQTLTYHPPHTPLITA--NTPPDQLLTPHYWTQQARNTVDYATTTQT 2895
Db 3914 PHVELIRDELLDITSDSSSQTPLVPWLSTVDGTWVDSPLDGEYWYRNLREPVGFHPAVSQ 3973
Qy 2896 LHQHGVTTYIELG---------PDNTLTTLTHHNLPNPPTTTLT-LTHPHHHPQTHLLTN 2945
Db 3974 LQAQGDTVFVEVSASPVLLQAMDDDVVTVATLRRDDGDATRMLTALAQAYVHGVTVDWPA 4033
Qy 2946 LAKTTTTWHPHHYTHHDNQPHTHTHLDLPTYPFQHHHYWLESTQPGAGNVSAAGLDPTEH 3005
Db 4034 ILGTTTT----------------RVLDLPTYAFQHQRYWLKSVDRAAAD---------GH 4068
Qy 3006 PLLGATLELATDGGALLAGRLSLRSHPWLADHAVGGTVLLSGATFLELALHAGTYVGCDR 3065
Db 4069 PLLGTVVALPGSDGVVLTGRVSLATHTWLADHAVRGSVLLPGTAFVELVVRAADEVECDV 4128
Qy 3066 VDELTLHAPLVVPVDGGVSVQVGVAAADGEGRRLVSVYARGGSACGGGGASGGVWTCHAS 3125
Db 4129 VDELVIETPLLLPQTGGVQLSVSVAEADESGHRTVTVFSQADNT--------DAWIRHVS 4180
Qy 3126 GVLVEAAAGGVVVDGLAGVWPPRGAVAVDVDGVRDRLAGAGCVLGPVFSGLRAVWRDGGD 3185
Db 4181 ATISTSDTPLSLPE--SDTWPPAQAQPMNVADFYDRLMAEGYEYGPAFQGLQAAWRDGDT 4238
Qy 3186 LLAEVCLPEEAWGDAAGFGLHPALLDGVVQPLSVLLPGGTGFGEGAGFGEGVRVPAVWGG 3245
Db 4239 VYAEVALAEEQAQEAARYDVHPALLDAALHACTLNTPDAE---------EGVGLPFSWNH 4289
Qy 3246 VSLHRAGVTGVRVRVSAVGRGGGREAVSVVVGDEAGVPVASVDRLELRPVDMGQLRAVSV 3305
Db 4290 VQVHATGPATLRVAVTQEADGW-----SLRVADDTGRPVATIGSLITRPVTADTLGSTT- 4343
Qy 3306 SAGRRGSLYAVQWAEVGPVPVCGQAWAWHEDVGESGGGPVPGVVVLRC-PDAGAGGGGGG 3364
Db 4344 -----DDLLTLTWTVIPPPQETSLPMVRYEDLAD-GNMPVPDVVVFTALPNSSENPLGPR 4397
Qy 3365 GGGG----GVGEVVGGVLGVVQGWLGLERFAGSRLVVVTRGAVVAGPEDGPVDVVGASVW 3420
Db 4398 PAGPADPLAQPDTDRQVLKPFQTWLP-GTFTDSTLVVRTGTGLAA----------AAVS 4445
Qy 3421 GLVRSAQAEHPDRFVLLDLDTDTGTDLDTGAGAGWGVDGGRVAAVVACGEPQLAVRGERL 3480
Db 4446 GLMRSAQSEHPGRFILVESDDDTLT----------LRQLAATVGLDEPRLRINDGRY 4492
Qy 3481 LAARLKRLESS-GDVP----------AQRSG----------DTRAR-------RSDV- 3509
Db 4493 EAARLARTNTTPEDTPLTIPDSRAWLLEQPRSGTLQDLALVPTDTAEQPLRPGEVRIDVR 4552
Qy 3510 ----------PAQRSGGV------PA--- 3519
Db 4553 AAGLNFRDVVVALGMVNDNRPTGGEAAGVVLEVGPDVQDLAPGDRVFGVVGGSFGPAAIA 4612

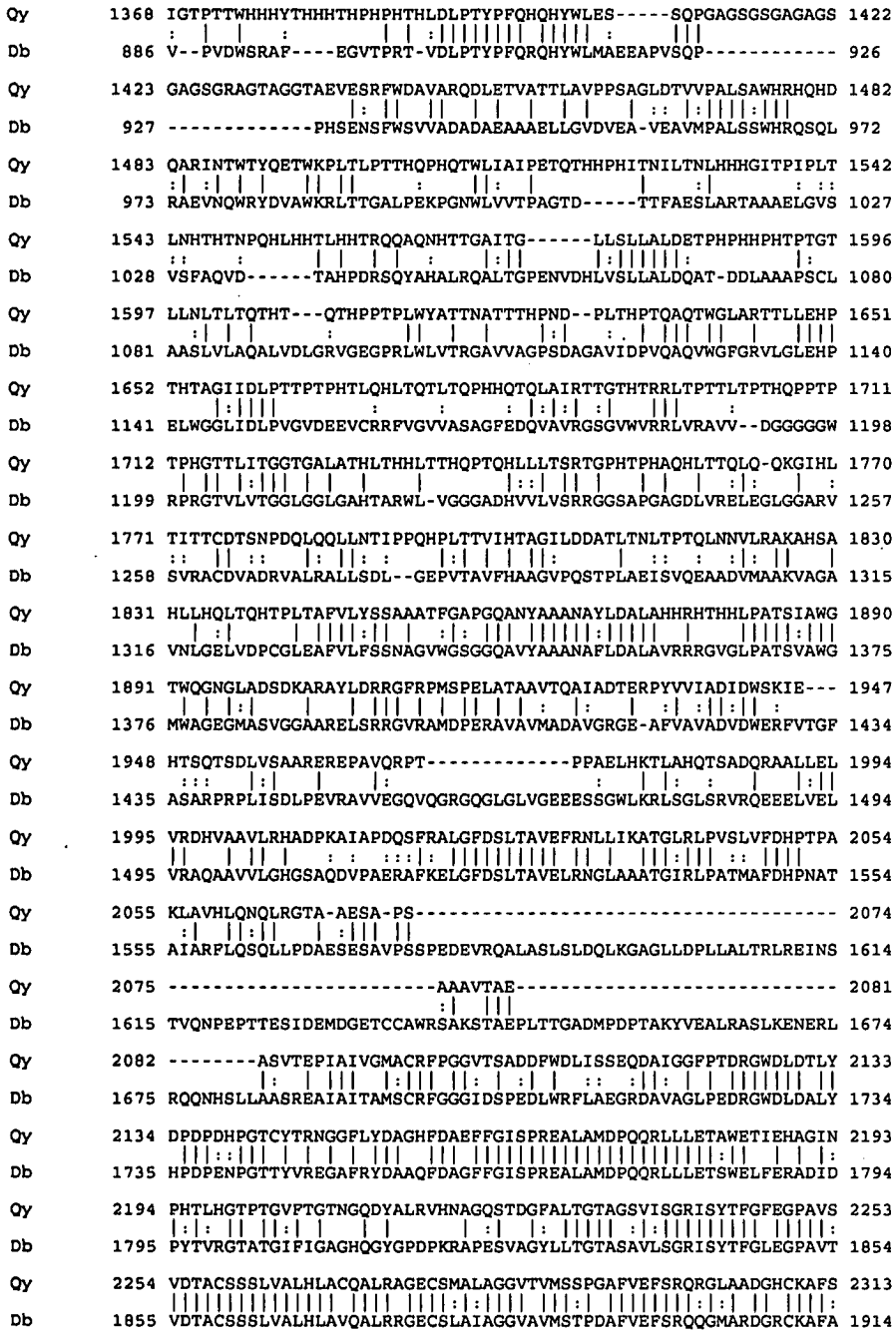


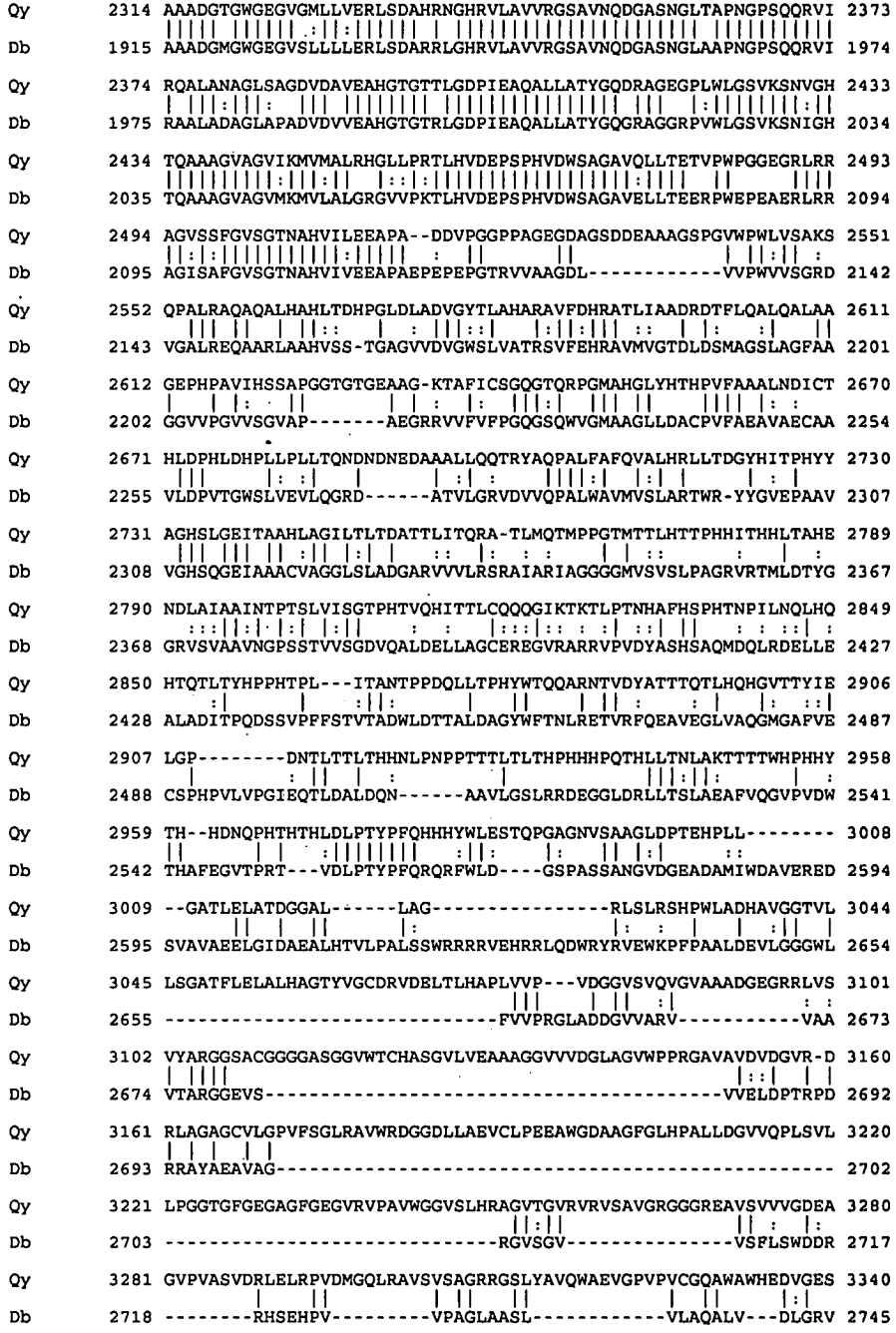
RESULT 11
S43048
polyketide synthase type I - Streptomyces antibioticus
N;Contains: acyl carrier protein; acyltransferase; ketoacylsynthase; ketoreductase; thio
C;Species: Streptomyces antibioticus
C;Date: 13-Jan-1995 #sequence_revision 06-Dec-1996 #text_change 26-May-2000
C;Accession: S43048; S41729
R;Swan, D.G.; Rodriguez, A.M.; Vilches, C.; Mendez, C.; Salas, J.A.
submitted to the EMBL Data Library, February 1993
A;Reference number: S43048
A;Accession: S43048
A;Molecule type: DNA
A;Residues: 1-3519 <SWA>
A;Cross-references: EMBL:L09654; NID:g153407; PIDN:AAA19695.1; PID:g153408
R;Swan, D.G.; Rodriguez, A.M.; Vilches, C.; Mendez, C.; Salas, J.A.
Mol. Gen. Genet. 242, 358-362, 1994
A;Title: Characterisation of a Streptomyces antibioticus gene encoding a type I polyketi
A;Reference number: S41729; MUID:94150470; PMID:8107683
A;Accession: S41729
A;Status: nucleic acid sequence not shown
A;Molecule type: DNA
A;Residues: 1683-3238,3273-3303,'K',3305-3407,'T',3409-3462,'Y',3464-3516,'E',3518-3519
A;Cross-references: EMBL:L09654
C;Genetics:
A;Start codon: GTG
C;Superfamily: [acyl-carrier-protein] S-malonyltransferase homology; 3-oxoacyl-[acyl-car
ogy; short-chain alcohol dehydrogenase homology
C;Keywords: antibiotic biosynthesis; carrier protein

F;59-462/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS1>
F;570-851/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT1>
F;1202-1381/Domain: short-chain alcohol dehydrogenase homology <SAD1>
F;1489-1560/Domain: acyl carrier protein homology <ACP1>
F;1708-2111/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS2>
F;2221-2502/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT2>
F;2858-3037/Domain: short-chain alcohol dehydrogenase homology <SAD2>
F;3143-3214/Domain: acyl carrier protein homology <ACP2>
F;3305-3500/Domain: oleoyl-[acyl-carrier-protein] hydrolase homology <ACPH>

Query Match 29.3%; Score 6108.5; DB 2; Length 3519;
Best Local Similarity 41.2%; Pred. No. 6.2e-270;
Matches 1498; Conservative 361; Mismatches 1093; Indels 687; Gaps 81;







```
Db       2746 GEGP-------------------------------------------------RLWLVTR 2756

Qy       3401 GAVVAGPED--GPVDVVGASVWGLVRSAQAEHPDRF-VLLDLDTD----------TGTDL 3447
Db       2757 DAVVAGPSDAGAVIDPVQAQVWGFGRVLGLEHPELWGGLIDLPVEAPEPGSTCDHTYADL 2816

Qy       3448 DTGAGAGWGVDGGRVAAVVACG--EPQLAVRGERLLAARLKRLESSGDVPAQRSGDTRAR 3505
Db       2817 --------------LATVVASAGFEDQVAVRGSGVWVRRLVRAVVDGG------------ 2850

Qy       3506 RSDVPAQRSGGVPARRSVDVSGREVLPWLSGGSVLVTGGTGVLGAAVARHLAGVCGVRDL 3565
Db       2851 --------GGG----------------WRPRGTVLVTGGLGGLGAHTARWLVG-GGADHV 2885

Qy       3566 LLVSRRGPDAPGAEGLRAELAAL-GAEVRIVACDVGERREVVRLLEGVPAGCPLTGVVHA 3624
Db       2886 VLVSRRGGSAPGAGDLVRELEGLGGARVSVRACDVADRVALRALLSDL--GEPVTAVFHA 2943

Qy       3625 AGVLDDATIASLTPERLGTVFAAKVDAALLLDELTRGMELSAFVLFSSAAGILGSAGQGN 3684
Db       2944 AGVPQSTPLAEISVQEAADVMAAKVAGAVNLGELVDPCGLEAFVLFSSNAGVWGSGGQAV 3003

Qy       3685 YAAANAALDALAYRRRAAGLPGVSLAWGLWEEASGMTGHLAGTDHRRIIRSGLHPMSTPD 3744
Db       3004 YAAANAFLDALAVRRRGVGLPATSVAWGMW-AGEGMAS--VGGAARELSRRGVRAMDPER 3060

Qy       3745 ALALFDAALALDRPVLLPADL------------RPAPPLPPLLQDLLPATR-----RRTT 3787
Db       3061 AVAVMADAVGRGEAFVAVADVDWERFVTGFASARPR----PLISD-LPEVRTALRNQEQE 3115

Qy       3788 RTTTTGGADNGAQLHARLAGQTHEQQHTTLLALVRSHIATVLGHTTPDTIPPDRAFRDLG 3847
Db       3116 QLHAPVPEDRSAQLLRRLSMLSPAGREAELVKLVRTEAAAVLGHGSAQDVPAERAFKELG 3175

Qy       3848 FDSLTAVELRNRLSRTTGLRLPTTLAFDHPNPTTLTHHL 3886
Db       3176 FDSLTAVQLRNRLAAATGTRLPASAVFDHPHAAALARWL 3214


RESULT 12
T17467
rifamycin polyketide synthase modules 9-10 - Amycolatopsis mediterranei
C;Species: Amycolatopsis mediterranei
C;Date: 02-Sep-2000 #sequence_revision 02-Sep-2000 #text_change 17-Nov-2000
C;Accession: T17467
R;Schupp, T.
submitted to the EMBL Data Library, December 1997
A;Reference number: Z18802
A;Accession: T17467
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-3413 <SCH>
A;Cross-references: EMBL:AJ223012; NID:e1227119; PID:e1227124; PIDN:CAA11039.1
A;Experimental source: strain LBG A3136
C;Superfamily: acyl carrier protein homology
C;Keywords: carrier protein
F;1608-1679/Domain: acyl carrier protein homology <ACP1>
F;3334-3405/Domain: acyl carrier protein homology <ACP2>

  Query Match           29.2%;  Score 6077.5;  DB 2;  Length 3413;
  Best Local Similarity  41.0%;  Pred. No. 1.5e-268;
  Matches 1507;  Conservative  408;  Mismatches 1210;  Indels  553;  Gaps   92;

Qy        475 GTAAADDPIAIIGMACRFPGGVRSPKDLWELAASGGDAIGPFPTDRGWPTEQRHAQDPTQ 534
Db         25 GARHADEPLAVVGMACRFPGGVSSPEDLWQLVAGGVDALSDFPDDRGWELDGLFDPDPDH 84

Qy        535 PGTFYPQGGGFLHDAAHFDAGFFGISPREALAMDPQQRLLLETSWEAFERAGIDPLSVRG 594
Db         85 PGTSYTSQGGFLRGAGLFDAGLFGISPREALVMDPQQRVLLETSWEALEDAGVDPLSLKG 144

Qy        595 SRTGVFAGALSFDYGPRMDTASSEGAADVEGHILTGTTGSVLSGRIAYSFGLEGPAITVD 654

Db        145 SDVGVFSGVFTQGYGAGAITPDLEAFAGI------GAASSVASGRVSYVFGLEGPAVTID 198

Qy        655 TGCSASLVTLHLACQSLRSGECTLALAGGVSVMSTLGMFIEFSRQRGLSVDGRCKAYSAA 714
Db        199 TACSSSLVAIHLAAQALRAGECSMALAGGATVMPTPGTFVAFSRQRVLAADGRSKAFSST 258

Qy        715 ADGTGWGEGVGMLLVERLSDAVRLGHRVLAVVRGSAVNQDGASNGLTAPNGPAQERVIRQ 774
Db        259 ADGTGWAEGAGVLVLERLSVAQERGHRILAVLRGSAVNQDGASNGLTAPNGPSQQRVIRK 318

Qy        775 ALANAGLSVADVDVVEGHGTGTTLGDPIEAQALLATYGQRAGDRPLWLGSLKSNIGHTMA 834
Db        319 ALAGAGLVASDVDVVEAHGTGTALGDPIEAQALLATYGQ-GRERPLWLGSVKSNFGHTQA 377

Qy        835 AAGVGGVIKMVMALREGVLPRTLHVDKPSPQVDWSAGAVRLLTEAVPWPGDAAGRLRRAG 894
Db        378 AAGVAGVIKMVQALRHGAMPPTLHVAEPTPEVDWSAGAVELLTEPREWP--AGDRPRRAG 435

Qy        895 VSSFGIGGTNAHVILEEAPAAGGCVAGGGVLEGAPGLAISVAESVAAPVAVSAPVAESVP 954
Db        436 VSAFGISGTNAHLILEEAPPAD-----------------AVAEE---------PEFKG-- 467

Qy        955 VPVPVPVPVPVSARSEAGLRAQAEALRQYVAVRPDVSLADVGAGLACGRAVLEHRAVVLA 1014
Db        468 -----PVPLVVSAGSPTSLAAQAGRLAEVLA-SGGVSRARLASGLLSGRALLGDRAVVVA 521

Qy       1015 ADREELVQGLGALAAGEPDRRVTTGHAPGGDRGGVVFVFPGQGGQWAGMGVRLLASSPVF 1074
Db        522 GTDEDAVAGLRALARGDRAPGVLTGSA---KHGKVVYVFPGQGSQRLGMGRELYDRYPVF 578

Qy       1075 ARRM-QACEE---ALAPWVDWSVVDILRRD--AGDAVWERADVVQPVLFSVMVSLAALWR 1128
Db        579 ATAFDEACEQLDVCLAGRAGHRVRDVVLGEVPAETGLLNQTVFTQAGLFAVESALFRLAE 638

Qy       1129 SYGIEPDAVLGHSQGEIAAAHVCGALSLKDAAKTVALRSRALAAVRGRGGMASVPLPAQE 1188
Db        639 SWGVRPDVVLGHSIGEITAAYAAGVFSLPDAARIVAARGRLMQALAPGGAMVAVAASEAE 698

Qy       1189 VEQLIGERWAGRLWVAAVNGPRSTAVSGDAEAVDEVLAYCAGTGVRARRIPVDYASHCPH 1248
Db        699 VAELLGD----GVELAAVNGPSAVVLSGDADAVVAAAARMRERGHKTKQLKVSHAFHSAR 754

Qy       1249 VQPLREELLELLGDISPQPSGVPFFSTVEGTWLDTTTL-DAAYWYRNLHQPVRFSDAVQA 1307
Db        755 MAPMLAEFAAELAGVTWREPEIPVVSNVTGRFAEPGELTEPGYWAEHVRRPVRFAEGVAA 814

Qy       1308 LADDGHRVFVEVSPHPTLVPAIEDTTEDTAEDVTAIGSLRRGDNDTRRFLTALAHTHTTG 1367
Db        815 ATESGGGSLFVELGPGAALTALVEETAE-----VTCVAALRDDRPEVTALITAVAELFVRG 869

Qy       1368 IGT-------PTTWHHHYTHHHTHPHPHTHLDLPTYPFQHQHYWLESSQPGAGSGSGAGA 1420
Db        870 VAVDWPALLPPVT--------------GFVDLPKYAFDQQHYWL---QPAAQATDAASL 911

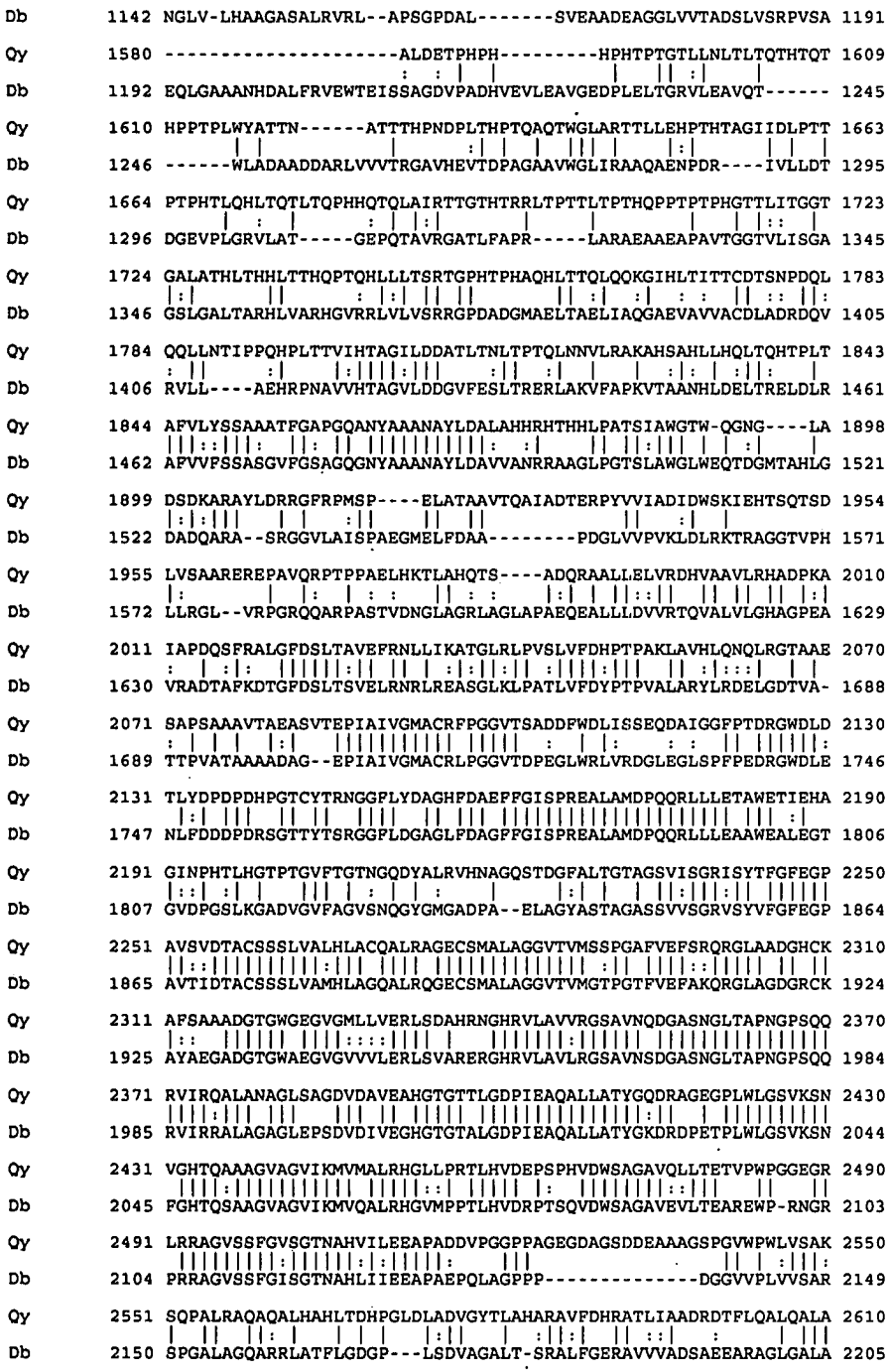
Qy       1421 GSGAG-----------------------------------SGRAGTAG-GTAEVESRFW 1443
Db        912 GQVAADHPLLGAVVRLPQSDGLVFTSRLSLKSHPWLADHVIGGVVLVAGTGLVELAVRAG 971

Qy       1444 DAVARQDLE--TVATTLAVPPSAG--LDTVVPA------------------LSAWHRH- 1479
Db        972 DEAGCPVLEELVIEAPLVVPDHGGVRIQVVVGAPGETGSRAVEVYSLREDAGAEVWARHA 1031

Qy       1480 ---------QHDQARINTW----------------------TYQETWKPLT--------- 1499
Db       1032 TGFLAATPSQHKPFDFTAWPPPGVERVDVEDFYDGLVDRGYAYGPSFRGLRAVWRRGDEV 1091

Qy       1500 ---------------------------------LPTTHQPHQTWLIAIPETQTHHPHIT 1525
Db       1092 FAEVALAEDDRADAARFGIHPGLLDAALHAGMAGATTTEEPGRPVL----------PFAW 1141

Qy       1526 NILTNLHHHGITPIPLTLNHTHTNPQHLHHTLHHTRQQAQNHTTGAITGLLSLL------ 1579
```


Db 1142 NGLV-LHAAGASALRVRL--APSGPDAL-------SVEAADEAGGLVVTADSLVSRPVSA 1191
Qy 1580 ----------ALDETPHPH---------HPHTPTGTLLNLTLTQTHTQT 1609
Db 1192 EQLGAAANHDALFRVEWTEISSAGDVPADHVEVLEAVGEDPLELTGRVLEAVQT------ 1245
Qy 1610 HPPTPLWYATTN------ATTTHPNDPLTHPTQAQTWGLARTTLLEHPTHTAGIIDLPTT 1663
Db 1246 ------WLADAADDARLVVVTRGAVHEVTDPAGAAVWGLIRAAQAENPDR----IVLLDT 1295
Qy 1664 PTPHTLQHLTQTLTQPHHQTQLAIRTTGTHTRRLTPTTLTPTHQPPTPTPHGTTLITGGT 1723
Db 1296 DGEVPLGRVLAT-----GEPQTAVRGATLFAPR-----LARAEAAEAPAVTGGTVLISGA 1345
Qy 1724 GALATHLTHHLTTHQPTQHLLLTSRTGPHTPHAQHLTTQLQQKGIHLTITTCDTSNPDQL 1783
Db 1346 GSLGALTARHLVARHGVRRLVLVSRRGPDADGMAELTAELIAQGAEVAVVACDLADRDQV 1405
Qy 1784 QQLLNTIPPQHPLTTVIHTAGILDDATLTNLTPTQLNNVLRAKAHSAHLLHQLTQHTPLT 1843
Db 1406 RVLL----AEHRPNAVVHTAGVLDDGVFESLTRERLAKVFAPKVTAANHLDELTRELDLR 1461
Qy 1844 AFVLYSSAAATFGAPGQANYAAANAYLDALAHHRHTHHLPATSIAWGTW-QGNG----LA 1898
Db 1462 AFVVPSSASGVFGSAGQGNYAAANAYLDAVVANRRAAGLPGTSLAWGLWEQTDGMTAHLG 1521
Qy 1899 DSDKARAYLDRRGFRPMSP----ELATAAVTQAIADTERPYVVIADIDWSKIEHTSQTSD 1954
Db 1522 DADQARA--SRGGVLAISPAEGMELFDAA--------PDGLVVPVKLDLRKTRAGGTVPH 1571
Qy 1955 LVSAARAREPAVQRPTPPAELHKTLAHQTS----ADQRAALLELVRDHVAAVLRHADPKA 2010
Db 1572 LLRGL--VRPGRQQARPASTVDNGLAGRLAGLAPAEQEALLLDVVRTQVALVLGHAGPEA 1629
Qy 2011 IAPDQSFRALGFDSLTAVEFRNLLIKATGLRLPVSLVFDHPTPAKLAVHLQNQLRGTAAE 2070
Db 1630 VRADTAFKDTGFDSLTSVELRNRLREASGLKLPATLVFDYPTPVALARYLRDELGDTVA- 1688
Qy 2071 SAPSAAAVTAEASVTEPIAIVGMACRFPGGVTSADDFWDLISSEQDAIGGFPTDRGWDLD 2130
Db 1689 TTPVATAAAADAG--EPIAIVGMACRLPGGVTDPEGLWRLVRDGLEGLSPFPEDRGWDLE 1746
Qy 2131 TLYDPDPDHPGTCYTRNGGFLYDAGHFDAEFFGISPREALAMDPQQRLLLETAWETIEHA 2190
Db 1747 NLFDDDPDRSGTTYTSRGGFLDGAGLFDAGFFGISPREALAMDPQQRLLLEAAWEALEGT 1806
Qy 2191 GINPHTLHGTPTGVFTGTNGQDYALRVHNAGQSTDGFALTGTAGSVISGRISYTFGFEGP 2250
Db 1807 GVDPGSLKGADVGVFAGVSNQGYGMGADPA--ELAGYASTAGASSVVSGRVSYVFGFEGP 1864
Qy 2251 AVSVDTACSSSLVALHLACQALRAGECSMALAGGVTVMSSPGAFVEFSRQRGLAADGHCK 2310
Db 1865 AVTIDTACSSSLVAMHLAGQALRQGECSMALAGGVTVMGTPGTFVEFAKQRGLAGDGRCK 1924
Qy 2311 AFSAAADGTGWGEGVGMLLVERLSDAHRNGHRVLAVVRGSAVNQDGASNGLTAPNGPSQQ 2370
Db 1925 AYAEGADGTGWAEGVGVVVLERLSVARERGHRVLAVLRGSAVNSDGASNGLTAPNGPSQQ 1984
Qy 2371 RVIRQALANAGLSAGDVDAVEAHGTGTTLGDPIEAQALLATYGQDRAGEGPLWLGSVKSN 2430
Db 1985 RVIRRALAGAGLEPSDVDIVEGHGTGTALGDPIEAQALLATYGKDRDPETPLWLGSVKSN 2044
Qy 2431 VGHTQAAAGVAGVIKMVMALRHGLLPRTLHVDEPSPHVDWSAGAVQLLTETVPWPGGEGR 2490
Db 2045 FGHTQSAAGVAGVIKMVQALRHGVMPPTLHVDRPTSQVDWSAGAVEVLTEAREWP-RNGR 2103
Qy 2491 LRRAGVSSFGVSGTNAHVILEEAPADDVPGGPPAGEGDAGSDDEAAAGSPGVWPWLVSAK 2550
Db 2104 PRRAGVSSFGISGTNAHLIIEEAPAEPQLAGPPP----------DGGVVPLVVSAR 2149
Qy 2551 SQPALRAQAQALHAHLTDHPGLDLADVGYTLAHARAVFDHRATLIAADRDTFLQALQALA 2610
Db 2150 SPGALAGQARRLATFLGDGP---LSDVAGALT-SRALFGERAVVVADSAEEARAGLGALA 2205
Qy 2611 AGEPHPAVIHSSAPGGTGTGEAAGKTAFICSGQGTQRPGMAHGLYHTHPVFAAALNDICT 2670
Db 2206 RGEDAPGLVRGRVP----ASGLPGKLVWVFPGQGTQWVGMGRELLEESPVFAERIAECAA 2261
Qy 2671 HLDPHLDHPLLPLLTQNDNDNEDAAALLQQTRYAQPALFAFQVALHRLLTDGYHITPHYY 2730
Db 2262 ALEPWIGWSLFDVL-RGDGD-------LDRVDVLQPACFAVMVGLAAVWSSA-GVVPDAV 2312
Qy 2731 AGHSLGEITAAHLAGILTLTDATTLITQR--ATLMQTMPPGTMTTLHTTPHHITHHLTAH 2788
Db 2313 LGHSQGEIAAACVSGALSLEDAAKVVALRSQAIAAKLSGRGGMASVALGEADVVSRLA-- 2370
Qy 2789 ENDLAIAAINTPTSLVISGTPHTVQHITTLCQQQGIKTKTLPTNHAFHSPHTNPILNQLH 2848
Db 2371 -DGVEVAAVNGPASVVIAGDAQALDETLEALSGAGIRARRVAVDYASHTRHVEDIEDTLA 2429
Qy 2849 QHTQTLTYHPPHTPLITANT----PPDQLLTPHYWTQQARNTVDYATTTQTLHQHGVTTY 2904
Db 2430 EALAGIDARAPLVPFLSTLTGEWIRDEGVVDGGYWYRNLRGRVRFGPAVEALLAQGHGVF 2489
Qy 2905 IELGPDNTLTTLTHHNLPNPPT-----TTLTLTHPHHHPQ---THLLTNLAKT------- 2949
Db 2490 VELS--------AHPVLVQPITELTDETAAVVTGSLRRDDGGLRRLLTSMAELFVRGVEV 2541
Qy 2950 --TTTWHPHHYTHHDNQPHTHTHLDLPTYPFQHHHYWLESTQPGAGNVSAAGLDPTEHPL 3007
Db 2542 DWTSLVPP-------------ARADLPTYAFDHEHYWLRAADTASDAVS-LGLAGADHPL 2587
Qy 3008 LGATLELATDGGALLAGRLSLRSHPWLADHAVGGTVLLSGATFLELALHAGTYVGCDRVD 3067
Db 2588 LGAVVQLPQSDGLVFTSRLSLRSHPWLADHAVRDVVIVPGTGLVELAVRAGDEAGCPVLD 2647
Qy 3068 ELTLHAPLVVPVDGGVSVQVGVAAADGEGRRLVSVYARGGSACGGGGASGGVWTCHASGV 3127
Db 2648 ELVIEAPLVVPRRGGVRVQVALGGPADDGSRTVDVFSLREDA--------DSWLRHATGV 2699
Qy 3128 LV-EAAAGGVVVDGLAGVWPPRGAVAVDVDGVRDRLAGAGCVLGPVFSGLRAVWRDGG-- 3184
Db 2700 LVPENRPRGTAAFDFA-AWPPPEAKPVDLTGAYDVLADVGYGYGPTFRAVRAVWRRGSGN 2758
Qy 3185 --DLLAEVCLPEEAWGDAAGFGLHPALLDGVVQPLSVLLPGGTGFGEGAGFGEGVRVPAV 3242
Db 2759 TTETFAEIALPEDARAEAGRFGIHPALLDAALHSTMVSAAADT-----ESYGDEVRLPFA 2813
Qy 3243 WGGVSLHRAGVTGVRVRVSAVGRGGGREAVSVVVGDEAGVPVASVDRLELRPVDMGQLRA 3302
Db 2814 WNGLRLHAAGASVLRVRVAK----PERDSLSLEAVDESGGLVVTLDSLVGRPVSNDQLTT 2869
Qy 3303 VSVSAGRRGSLYAVQWAEVGPVPVCGQAWAW-----HEDVGE------SGGGPVPGVVVL 3351
Db 2870 AAGPAG-AGSLYRVDWTPLSSVDTSGRVPSWLPVATAEEVATLADDVLTGATEAPAVAVM 2928
Qy 3352 RCPDAGAGGGGGGGGGGVGEVVGGVLGVVQGWL---GLERFAGSRLVVVTRGAVVAGPE 3408
Db 2929 EAV----------ADEGSVLALTVRVLDVVQCWLAGGGLE---GTKLAIVTRGAVPAG-- 2973
Qy 3409 DGPV-DVVGASVWGLVRSAQAEHPDRFVLLDLDTDTGTDLDTGAGAGWGVDGGRVAAVVA 3467
Db 2974 DGVVHDPAAAAVWGLVRAAQAENPDRIVLLDVEPE--ADVPPLLG----------SVLA 3020
Qy 3468 CGEPQLAVRGERLLAARLKRLESSGDVPAQRSGDTRARRSDVPAQRSGGVPARRSVDVSG 3527
Db 3021 DGEPQVAVRGTTLSIPRL---------------ARAARPD-PA---AGFKTR------- 3053
Qy 3528 REVLPWLSGGSVLVTGGTGVLGAAVARHLAGVCGVRDLLLVSRRGPDAPGAEGLRAELAA 3587
Db 3054 ---------GPVLVTGGTGSLGGLVARHLVERHGVRQLVLASRRGLDAEGAKDLVTDLTA 3104
Qy 3588 LGAEVRIVACDVGERREVVRLL-EGVPAGCPLTGVVHAAGVLDDATIASLTPERLGTVFA 3646
Db 3105 LGADVAVAACDVADRDQVAALLTEHRP-----SAVVHTAGVPDAGVIGTVTPDRLAEVFA 3159

```
Qy       3647 AKVDAALLLDELTRGMELSAFVLFSSAAGILGSAGQGNYAAANAALDALAYRRRAAGLPG 3706
Db       3160 PKVTAARHLDELTRDLDLDSFVVYSSVSAVFMGAGSGSYAAANAYLDGLMAHRRAAGLPG 3219

Qy       3707 VSLAWGLWEEASGMTGHLAGTDH----RRIIRSGLHPMSTPDALALFDAALALDRPVLLP 3762
Db       3220 QSLAWGLWDQTTG--GMAAGTDEAGRARMTRRGGLVAMKPAAGLDLFDAAIGSGEPLLVP 3277

Qy       3763 A--DLR-------PAPPLPPLLQDLLPATRRRTTRTTTTGGADNGAQLHARLAGQTHEQQ 3813
Db       3278 AQLDLRGLRAEAAGGTEVPHLLRGLVRAGRQQARAASTV--EENWA---GRLAGLEPAER 3332

Qy       3814 HTTLLALVRSHIATVLGHTTPDTIPPDRAFRDLGFDSLTAVELRNRLSRTTGLRLPTTLA 3873
Db       3333 GQVLLELVRAQVAGVLGYRAAHQVDPDQGLFEIGFDSLTAIELRNRLRARTERKISPGVV 3392

Qy       3874 FDHPNPTTLTHHLHTQLQ 3891
Db       3393 FDHPTPALLAAHLNELLR 3410


RESULT 13
T17463
rifamycin polyketide synthase modules 1-3 - Amycolatopsis mediterranei
C;Species: Amycolatopsis mediterranei
C;Date: 02-Sep-2000 #sequence_revision 02-Sep-2000 #text_change 17-Nov-2000
C;Accession: T17463
R;Schupp, T.
submitted to the EMBL Data Library, December 1997
A;Reference number: Z18802
A;Accession: T17463
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-4735 <SCH>
A;Cross-references: EMBL:AJ223012; NID:e1227119; PID:e1227120; PIDN:CAA11035.1
A;Experimental source: strain LBG A3136
C;Superfamily: acyl carrier protein homology; acetate-CoA ligase homology
C;Keywords: carrier protein
F;53-500/Domain: acetate-CoA ligase homology <ACL>
F;543-610/Domain: acyl carrier protein homology <ACP1>
F;2102-2173/Domain: acyl carrier protein homology <ACP2>
F;3079-3150/Domain: acyl carrier protein homology <ACP3>
F;4578-4649/Domain: acyl carrier protein homology <ACP4>

  Query Match             29.1%;  Score 6068.5;  DB 2;  Length 4735;
  Best Local Similarity   35.2%;  Pred. No. 6.1e-268;
  Matches 1745;  Conservative  422;  Mismatches 1437;  Indels 1359;  Gaps  143;

Qy         37 PQWPGMGRELLD-------ASDVFRESVRACEAAFAPYVD------------WSVEQVLR 77
Db         86 PLNPGSATAELEHPLTDSGATVVVTDAAQAARLRLAPHVELLVTGDDVPEGAHSYDELAL 145

Qy         78 DSPDAPGLDRVDVVQPT-LF-----------AVMISLAALWRSQGVEPCAV--------- 116
Db        146 SEPAEPAADDLELDEPAWMFYTSGTTGRPKGVVSTQRNCLW---SVASCYVPFPGLSDQD 202

Qy        117 -------LGHSLGEI----------AAAHVSGGLSLADAARVVTLWSQAQTTLAG----- 154
Db        203 RVLWPLPLFHSLSHIACVLSATVVGASVRIADGSSADDVMRLIE--AESSTFLAGVPTTY 260

Qy        155 --------------------TGALVSVAATPDEL-----LPRIAPWTEDNPARLAVAAV 188
Db        261 HHLVRAARQRGFSAPSLRIGLAGGAVLGAGLRSEFEETFGVPLIDAY--GSTETCGAITM 318

Qy        189 NGPRSTVVSGAREAVADLVADLTAAQVRTRMIP----VDVPA----------------- 226
Db        319 NPPDGARVEGS--------CGLAVPGVDVRVVDPDTGLDVPAGEEGEVWVSGPNVMLGYH 370

Qy        227 HSPLMYA-------------------------IEERVVSGLLPITPRPSRIPFH---- 255
Db        371 NSPEATAAAMRDGWFRTGDLARRDDAGYFTICGRIKELIIRGGANIHPGEVEAVLRTVDG 430

Qy        256 ---SSVTGGRLDT-RELDAAYWYRNMSSTVRFEPAARLLLQQGPKTFVEMSPHPVLTMGL 311
```

```
Db        431 VADAAVGGVPHDTLGEVPVAY---VIPGPTGFDPAALI-----EKCREQLSAYKV----- 477

Qy        312 QELAPDLGDTTGTADTVIMGTLRRGQGT-----LDHFLTSLAQLRGHGETSATTVLSARL 366
Db        478 ----PDRILEVAHIPRTASGKIRRGLLTDEPAQLRYAATEHEEQSRHADESVAAALRARL 533

Qy        367 TALSPTQQQSLLLDLVRAHTMAVLND---DGNERTASDAGPSASFAHLGFDSVMGVELRN 423
Db        534 SGLDERAQCELLEDLVRTQAADVLGQPVPDGR-----------AFRDLGFTSLAIVELRN 582

Qy        424 RLSKATGLRLPVTLIFDHTTPAAVAARLRTAALGHLDEDTAPVPDSPSGHGGTAAAD--D 481
Db        583 RLTEHTGLWLPASAVFDHPTPAALAARVRAELLGITQAVAEPV----------VAADPGE 632

Qy        482 PIAIIGMACRFPGGVRSPKDLWELAASGGDAIGPFPTDRGWPTEQRHAQDPTQPGTFYPQ 541
Db        633 PIAIVGMACRLPGGVASPEDLWRLVAERVDAVSEFPGDRGWDLDSLIDPDRERAGTSYVG 692

Qy        542 GGGFLHDAAHFDAGFFGISPREALAMDPQQRLLLETSWEAFERAGIDPLSVRGSRTGVFA 601
Db        693 QGGFLHDAGEFDAGFFGISPREAVAMDPQQRLLLETSWEALENAGVDPIALKGTDTGVFS 752

Qy        602 GALSFDYGPRMDTASSEGAADVEGHILTGTTGSVLSGRIAYSFGLEGPAITVDTGCSASL 661
Db        753 GLMGQGYG------SGAVAPELEGFVTTGVASSVASGRVSYVLGLEGPAVTVDTACSSSL 806

Qy        662 VTLHLACQSLRSGECTLALAGGVSVMSTLGMFIEFSRQRGLSVDGRCKAYSAAADGTGWG 721
Db        807 VAMHLAAQALRQGECSMALAGGVTVMATPGSFVEFSRQRALAPDGRCKAFAAAADGTGWS 866

Qy        722 EGVGMLLVERLSDAVRLGHRVLAVVRGSAVNQDGASNGLTAPNGPAQERVIRQALANAGL 781
Db        867 EGVGVVVLERLSVARERGHRILAVLRGSAVNQDGASNGLTAPNGLSQQRVIRRALAAAGL 926

Qy        782 SVADVDVVEGHGTGTTLGDPIEAQALLATYGQRAGDRPLWLGSLKSNIGHTMAAAGVGGV 841
Db        927 APSDVDVVEAHGTGTTLGDPIEAQALLATYGQER-KQPLWLGSLKSNIGHAQAAAGVAGV 985

Qy        842 IKMVMALREGVLPRTLHVDKPSPQVDWSAGAVRLLTEAVPWPGDAAGRLRRAGVSSFGIG 901
Db        986 IKMVQALRHETLPPTLHVDKPTLEVDWSAGAIELLTEARAWPRN--GRPRRAGVSSFGVS 1043

Qy        902 GTNAHVILEEAPAAGGCVAGGGVLEGAPGLAISVAESVAAPVAVSAPVAESVPVPVPVPV 961
Db       1044 GTNAHLILEEAPA----------------------EEPVAAP---------ELPV-----V 1068

Qy        962 PVPVSARSEAGLRAQAEALRQYVAVRPDVSLADVGAGLACGRAVLEHRAVVLAADREELV 1021
Db       1069 PLVVSARSTESLSGQAERLAS--LLEGDVSLTEVAGALVSRRAVLDERAVVVAGSREEAV 1126

Qy       1022 QGLGALAAGEPDRRVTTGHAPGGDRGGVVFVFPGQGGQWAGMGVRLLASSPVFARRMQAC 1081
Db       1127 TGLRAL-----------NTAGSGTPGKVVWVFPGQGTQWAGMGRELLAESPVFAERIAEC 1175

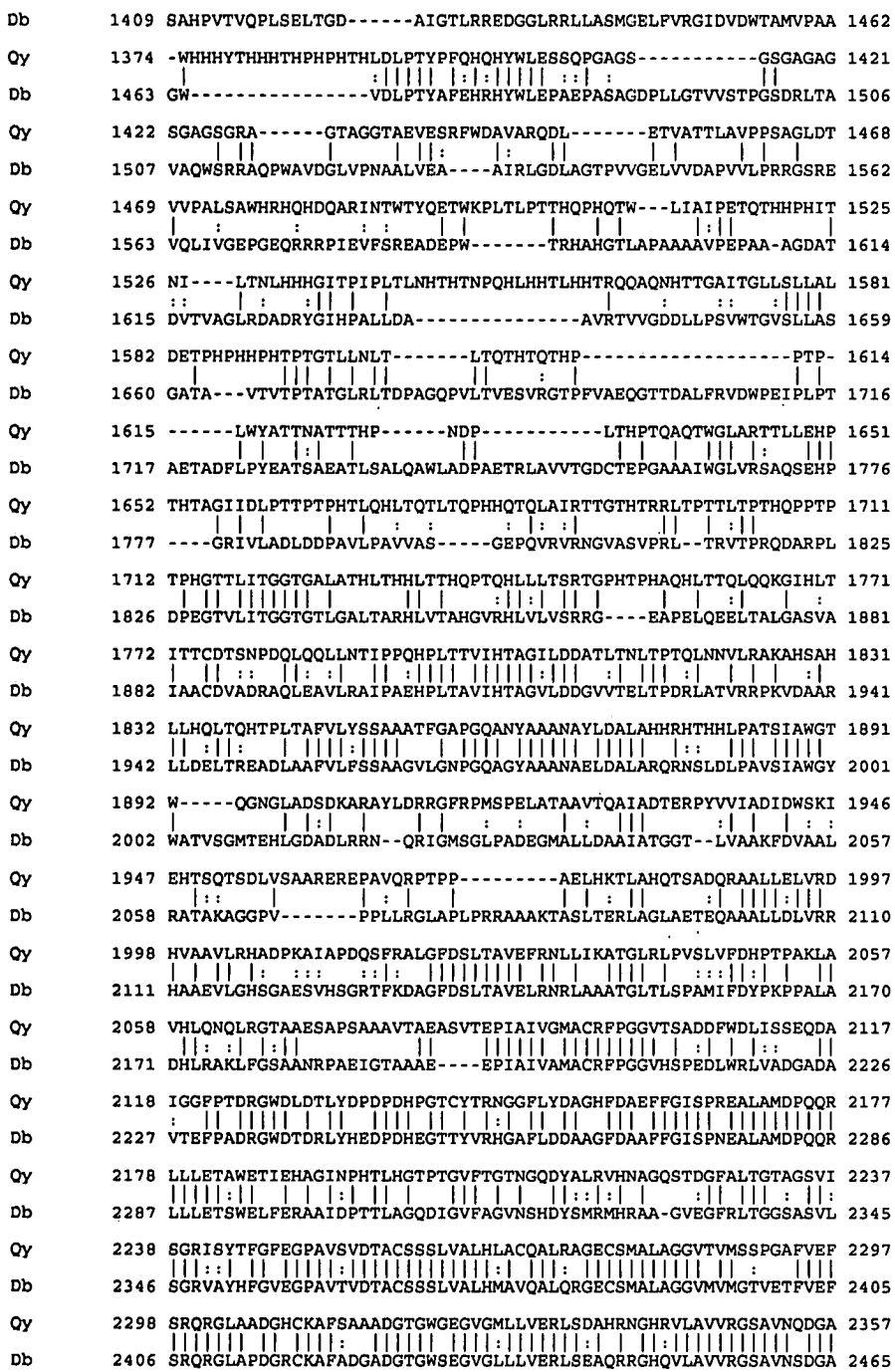
Qy       1082 EEALAPWVDWSVVDILRRDAGDAVWERADVVQPVLFSVMVSLAALWRSYGIEPDAVLGHS 1141
Db       1176 AAALAPWIDWSLVDVLR---GEGDLGRVDVLQPACFAVMVGLAAVWESVGVRPDAVVGHS 1232

Qy       1142 QGEIAAAHVCGALSLKDAAKTVALRSRALAA-VRGRGGMASVPLPAQEVEQLIGERWAGR 1200
Db       1233 QGEIAAACVSGALSLEDAAKVVALRSQAIAAELSGRGGMASVALGEDDVV----SRLVDG 1288

Qy       1201 LWVAAVNGPRSTAVSGDAEAVDEVLAYCAGTGVRARRIPVDYASHCPHVQPLREELLELL 1260
Db       1289 VEVAAVNGPSSVVIAGDAHALDATLEILSGEGIRVRRVAVDYASHTRHVEDIRDTLAETL 1348

Qy       1261 GDISPQPSGVPFFSTVEGTWL-DTTTLDAAYWYRNLHQPVRFSDAVQALADDGHRVFVEV 1319
Db       1349 AGISAQAPAVPFYSTVTSEWVRDAGVLDGGYWYRNLRNQVRFGAAATALLEQGHTVFVEV 1408

Qy       1320 SPHPTLVPAIEDTTEDTAEDVTAIGSLRRGDNDTRRFLTALAHTHTTGIGTPTT------ 1373
```


Db 1409 SAHPVTVQPLSELTGD------AIGTLRREDGGLRRLLASMGELFVRGIDVDWTAMVPAA 1462
Qy 1374 -WHHHYTHHHTHPHPHTHLDLPTYPFQHQHYWLESSQPGAGS----------GSGAGAG 1421
Db 1463 GW---------------VDLPTYAFEHRHYWLEPAEPASAGDPLLGTVVSTPGSDRLTA 1506
Qy 1422 SGAGSGRA------GTAGGTAEVESRFWDAVARQDL-------ETVATTLAVPPSAGLDT 1468
Db 1507 VAQWSRRAQPWAVDGLVPNAALVEA----AIRLGDLAGTPVVGELVVDAPVVLPRRGSRE 1562
Qy 1469 VVPALSAWHRHQHDQARINTWTYQETWKPLTLPTTHQPHQTW---LIAIPETQTHHPHIT 1525
Db 1563 VQLIVGEPGEQRRRPIEVFSREADEPW-------TRHAHGTLAPAAAAVPEPAA-AGDAT 1614
Qy 1526 NI----LTNLHHHGITPIPLTLNHTHTNPQHLHHTLHHTRQQAQNHTTGAITGLLSLLAL 1581
Db 1615 DVTVAGLRDADRYGIHPALLDA---------------AVRTVVGDDLLPSVWTGVSLLAS 1659
Qy 1582 DETPHPHHPHTPTGTLLNLT-------LTQTHTQTHP-------------------PTP- 1614
Db 1660 GATA---VTVTPTATGLRLTDPAGQPVLTVESVRGTPFVAEQGTTDALFRVDWPEIPLPT 1716
Qy 1615 ------LWYATTNATTTHP------NDP-----------LTHPTQAQTWGLARTTLLEHP 1651
Db 1717 AETADFLPYEATSAEATLSALQAWLADPAETRLAVVTGDCTEPGAAAIWGLVRSAQSEHP 1776
Qy 1652 THTAGIIDLPTTPTPHTLQHLTQTLTQPHHQTQLAIRTTGTHTRRLTPTTLTPTHQPPTP 1711
Db 1777 ----GRIVLADLDDPAVLPAVVAS-----GEPQVRVRNGVASVPRL--TRVTPRQDARPL 1825
Qy 1712 TPHGTTLITGGTGALATHLTHHLTTHQPTQHLLLTSRTGPHTPHAQHLTTQLQQKGIHLT 1771
Db 1826 DPEGTVLITGGTGTLGALTARHLVTAHGVRHLVLVSRRG----EAPELQEELTALGASVA 1881
Qy 1772 ITTCDTSNPDQLQQLLNTIPPQHPLTTVIHTAGILDDATLTNLTPTQLNNVLRAKAHSAH 1831
Db 1882 IAACDVADRAQLEAVLRAIPAEHPLTAVIHTAGVLDDGVVTELTPDRLATVRRPKVDAAR 1941
Qy 1832 LLHQLTQHTPLTAFVLYSSAAATFGAPGQANYAAANAYLDALAHHRHTHHLPATSIAWGT 1891
Db 1942 LLDELTREADLAAFVLFSSAAGVLGNPGQAGYAAANAELDALARQRNSLDLPAVSIAWGY 2001
Qy 1892 W-----QGNGLADSDKARAYLDRRGFRPMSPELATAAVTQAIADTERPYVVIADIDWSKI 1946
Db 2002 WATVSGMTEHLGDADLRRN--QRIGMSGLPADEGMALLDAAIATGGT--LVAAKFDVAAL 2057
Qy 1947 EHTSQTSDLVSAARAREPAVQRPTPP---------AELHKTLAHQTSADQRAALLELVRD 1997
Db 2058 RATAKAGGPV-------PPLLRGLAPLPRRAAAKTASLTERLAGLAETEQAAALLDLVRR 2110
Qy 1998 HVAAVLRHADPKAIAPDQSFRALGFDSLTAVEFRNLLIKATGLRLPVSLVFDHPTPAKLA 2057
Db 2111 HAAEVLGHSGAESVHSGRTFKDAGFDSLTAVELRNRLAAATGLTLSPAMIFDYPKPPALA 2170
Qy 2058 VHLQNQLRGTAAESAPSAAAVTAEASVTEPIAIVGMACRFPGGVTSADDFWDLISSEQDA 2117
Db 2171 DHLRAKLFGSAANRPAEIGTAAAE----EPIAIVAMACRFPGGVHSPEDLWRLVADGADA 2226
Qy 2118 IGGFPTDRGWDLDTLYDPDPDHPGTCYTRNGGFLYDAGHFDAEFFGISPREALAMDPQQR 2177
Db 2227 VTEFPADRGWDTDRLYHEDPDHEGTTYVRHGAFLDDAAGFDAAFFGISPNEALAMDPQQR 2286
Qy 2178 LLLETAWETIEHAGINPHTLHGTPTGVFTGTNGQDYALRVHNAGQSTDGFALTGTAGSVI 2237
Db 2287 LLLETSWELFERAAIDPTTLAGQDIGVFAGVNSHDYSMRMHRAA-GVEGFRLTGGSASVL 2345
Qy 2238 SGRISYTFGPEGPAVSVDTACSSSLVALHLACQALRAGECSMALAGGVTVMSSPGAFVEF 2297
Db 2346 SGRVAYHFGVEGPAVTVDTACSSSLVALHMAVQALQRGECSMALAGGVMVMGTVETFVEF 2405
Qy 2298 SRQRGLAADGHCKAFSAAADGTGWGEGVGMLLVERLSDAHRNGHRVLAVVRGSAVNQDGA 2357
Db 2406 SRQRGLAPDGRCKAFADGADGTGWSEGVGLLLVERLSEAQRRGHQVLAVVRGSAVNSDGA 2465
Qy 2358 SNGLTAPNGPSQQRVIRQALANAGLSAGDVDAVEAHGTGTTLGDPIEAQALLATYGQDRA 2417
Db 2466 SNGLTAPNGPSQQRVIRKALAAAGLSTSDVDAVEAHGTGTTLGDPIEAEALLATYGQNR- 2524
Qy 2418 GEGPLWLGSVKSNVGHTQAAAGVAGVIKMVMALRHGLLPRTLHVDEPSPHVDWSAGAVQL 2477
Db 2525 -ETPLWLGSVKSNLGHTQAAAGVAGVIKMVMAMRHGVLPRTLHVDRPSSYVDWSAGAVEL 2583
Qy 2478 LTETVPWPGGEGRLRRAGVSSFGVSGTNAHVILEEAPADDVPGGPPAGEGDAGSDDEAAA 2537
Db 2584 LTEARDWV-SNGHPRRAGVSSFGIGGTNAHVVLEEVAAPITTPQPEPAE---------- 2631
Qy 2538 GSPGVWPWLVSAKSQPALRAQAQALHAHLTDHPGLDLADVGYTLAHARAVFDHRATLIAA 2597
Db 2632 ---FLVPVLVSARTAAGLRGQAGRLAAFLGDRTDVRVPDAAYALATTRAQLDHRAVVLAS 2688
Qy 2598 DRDTFLQALQALAAGEPHPAVIHSSAPGGTGTGEAAGKTAFICSGQGTQRPGMAHGLYHT 2657
Db 2689 DRAQLCADLAAFGSG-----VV-------TGT-PVDGKLAVLFTGQGSQWAGMGRELAET 2735
Qy 2658 HPVFAAALNDICTHLDPHL-DHPLLPLLTQNDNDNEDAAALLQQTRYAQPALFAFQVALH 2716
Db 2736 FPVFRDAFEAACEAVDTHLRERPLREVVF-------DDSALLDQTMYTQGALFAVETALF 2788
Qy 2717 RLLTDGYHITPHYYAGHSLGEITAAHLAGILTLTDATTLITQRATLMQTMPP-GTMTTLH 2775
Db 2789 RLF-ESWGVRPGLLAGHSIGELAAAHVSGVLDLADAGELVAARGRLMQALPAGGAMVAVQ 2847
Qy 2776 TTPHHITHHLTAHENDLAIAAINTPTSLVISGTPHTVQHITTLCQQQGIKTKTLPTNHAF 2835
Db 2848 ATEDEVAPLL---DGTVCVAAVNGPDSVVLSGTEAAVLAVADELAGRGRKTRRLAVSHAF 2904
Qy 2836 HSPHTNPILNQLHQHTQTLTYHPPHTPLITANTPPDQLL-TPHYWTQQARNTVDYATTTQ 2894
Db 2905 HSPLMEPMLDDFRAVAERLTYRAGSLPVVSTLTGELAALDSPDYWVGQVRNAVRFSDAVT 2964
Qy 2895 TLHQHGVTTYIELGPDNTLTTLTHHNLPNPPTTTLTLTHPHHHPQTHLLTNLAKTTTTWH 2954
Db 2965 ALGAQGASTFLELGPGGALAAMALGTLGGPEQSCVATLRKNGAEVPDVLTALAELHVRGV 3024
Qy 2955 PHHYTHHDNQPHTHTHLDLPTYPFQHHHYWLESTQPGAGNVSAAGLDPTEHPLLGATLEL 3014
Db 3025 GVDWTTVLDEPATAVGTVLPTYAFQHQRFWVDVDETAAVSVTPPPAEPIVDRPVQDVLEL 3084
Qy 3015 ATDGGALLAGR---------LSLRSHPW------------------------------ 3033
Db 3085 VRESAAVVLGHRDAGSFDLDRSFKDHGFDSLSAVKLRNRLRDFTGVELPSTLIFDYPNPA 3144
Qy 3034 -LADH-------------------------AVGGTVLLSGATFLELAL---------- 3055
Db 3145 VLADHLRAELLGERPAAPAPVTRDVSDEPIAIVGMSTRLPGGADSPEELWKLVAEGRDAV 3204
Qy 3056 --------------------HAGT------------------------------------ 3059
Db 3205 SGFPVDRGWDLDGLYHPDPAHAGTSYTRSGGFLHDAAQFDAGLFGISPREALAMDPQQRL 3264
Qy 3060 -----YVGCDR--VDELT-----------LHAPLV-----VPVD-------------- 3080
Db 3265 LLETSWEALERAGVDPLSARGSDVGVFTGIVHHDYVTRLREVPEDVQGYTMTGTASSVAS 3324
Qy 3081 ---GGVSV---------- 3085
Db 3325 GRVAYVFGFEGPAVTVDTACSSSLVAMHLAAQALRQGECSMALAGGATVMASPDAFLEFS 3384
Qy 3086 -QVGVAA----------ADGEG----------RRLVSVYARG--------GSACGGGGAS 3116
Db 3385 RQRGLSADGRCKAYAEGADGTGWAEGVGVVVLERLSVARERGHRVLAVLRGSAVNQDGAS 3444
Qy 3117 GG--------------------------VWTCHASGVLV---------------------- 3129
Db 3445 NGLTAPNGPSQQRVIRGALASAGLAPSDVDVVEGHGTGTALGDPIEVQALLATYGQEREQ 3504

```
Qy       3130 --------------EAAAG--GVV-----------------------VDGLAGV------ 3144
                            :||||  ||:                      ||  ||
Db       3505 PLWLGSLKSNLGHTQAAAGVVGVIKMIMAMRHGVMPATLHVDERTSQVDWSAGAIEVLTE 3564

Qy       3145 ---WP----PR-------------------------------------------------- 3148
                 ||    ||
Db       3565 AREWPRTGRPRRAGVSSFGASGTNAHLIIEEGPAEEAVDEEVASVVPLVVSARSAGSLAG 3624

Qy       3149 ------------------GAV----------AVDVDGVRDR-------LA----GAGCVL 3169
                              ||:           || : | ||         ||      | |
Db       3625 QAGRLAAVLENESLAGVAGALVSGRATLNERAVVIAGSRDEAQDGLQALARGENAPGVVT 3684

Qy       3170 GP---------VFSGLRAVWRD-GGDLL----------AEVCLPEEAW----------GD 3199
                |         || |  : |   | |||           |     | |           ||
Db       3685 GTAGKPGKVVWVFPGQGSQWMGMGRDLLDSSPVFAARIKECAAALEQWTDWSLLDVLRGD 3744

Qy       3200 AAGFGLHPALLD--GVVQPLSVLLPGG-----TGFGE------GAGFGEGVRVPAVWGGV 3246
                |         |||   |||| |  :  |     |  |       |   || :    | | :
Db       3745 A-------DLLDRVDVVQPASFAMMVGLAAVWTSLGVTPDAVLGHSQGE-IAAACVSGAL 3796

Qy       3247 SLH-------------------RAGVTGV----------------RVRVSAVGRGGG--- 3268
                ||                  | |:  |               || |:||
Db       3797 SLDDAAKVVALRSQAIAGELAGRGGMASVALSEEDAVARLTPWANRVEVAAVNSPSSVVI 3856

Qy       3269 -------REAVSVVVGD------------------------------EAGVPV----- 3284
                     ||:  : ||                              :| ||
Db       3857 AGDAQALDEALEALAGDGVRVRRVAVDYASHTRHVEAIAETLAKTLAGIDARVPAIPFYS 3916

Qy       3285 ---------ASVD-----------------------------RLELRPVDMGQLRAVSV 3305
                         | ||                          :   || :  |  :|
Db       3917 TVLGTWIEQAVVDAGYWYRNLRQQVRFGPSVADLAGLGHTVFVEISAHPVLVQPLSEISD 3976

Qy       3306 SAGRRGSL---------------------YAVQWAEVGP---------------------- 3323
                |   |||                  || |    |
Db       3977 DAVVTGSLRRDDGGLRRLLASAAELYVRGVAVDWTAAVPAAGWVDLPTYAFDRRHFWLHE 4036

Qy       3324 ------------------------------------------VPVCGQAW--------- 3331
                                                        |||  | |
Db       4037 AETAEAAEGMDGEFWTAIEQSDVDSLAELLELVPEQRGALSTVVPVLAQ-WRDRRRERST 4095

Qy       3332 --------AWHEDVGESGGGPVPGVVVLRCPDAGAGGG------GGG------------- 3364
                       |     |: |  |||    |    ||          | |
Db       4096 AEKLRYQVTWQPLEREAAG--VPGGRWLAVVPAGTTDALLKELTGQGLDIVRLEIEEASR 4153

Qy       3365 -------------------------GGGGGVGEVVGGVLGVVQGWLGLERFAGSRLVVV 3398
                                       ||     |:    | :||  || :    : |  :
Db       4154 AQLAEQLRNVLAEHDLTGVLSLLALDGGPADAAEITASTLALVQA-LG-DTTTSAPLWCL 4211

Qy       3399 TRGAVVAGPEDGPVDVVGASVWGLVRSAQAEHPDRFVLLDLDTDTGTDLDTGAGAGWGVD 3458
                | |||  | :|       |:|||| |:    |  ||:  | :|      |  | | |  ||
Db       4212 TSGAVNIGIQDAVTAPAQAAVWGLGRAVALERLDRWGGL-VDLPAAIDART-AQALLGVL 4269

Qy       3459 GGRVAAVVACGEPQLAVRGERLLAARLKRLESSGDVPAQRSGDTRAR--RSDVPAQRSGG 3516
                |     | || |||||                    |||  | |  |  ||   :
Db       4270 NG------AAGEDQLAVR--------------------RSGVYRRRLVRKPVPESATS- 4301

Qy       3517 VPARRSVDVSGREVLPWLSGGSVLVTGGTGVLGAAVARHLAGVCGVRDLLLVSRRGPDAP 3576
                          |   |:||||||   ||   :  ||  |   |::   | |
Db       4302 --------------RWEPRGTVLVTGGAEGLGRHASVWLA-QSGAERLIVT---GTD-- 4340

Qy       3577 GAEGLRAELAALGAEVRIVACDVGERREVVRLLEGVPAGCPLTGVVHAAGVLDDATIASL 3636
                | : | |||| |  |     |  :|  : :|:    |   :| ||||| :  :::
Db       4341 GVDELTAELAEFGTTVEF--CADTDRDAIAQLV----ADSEVTAVVHAADIAQTSSVDDT 4394

Qy       3637 TPERLGTVFAAKVDAALLLDELTRGMELSAFVLFSSAAGILGSAGQGNYAAANAALDALA 3696
                  | ||||||  |: ||:|     | |||:||| ||| |  |||   |||| ||||
Db       4395 GVADLDEVFAAKVTTAVWLDQLFEDTPLDAFVVFSSIAGIWGGGGQGPAGAANAVLDALV 4454

Qy       3697 YRRRAAGLPGVSLAWGLWEEASGMTGHLAGTDH---RRIIRSGLHPMSTPDALALFDAAL 3753
                  ||| ||   |:|||  ::         | |     :: | |: ||: | |:     |:
Db       4455 EWRRARGLKATSIAWGALDQIG------IGMDEAALAQLRRRGVIPMAPPLAVTAMVQAV 4508

Qy       3754 ALDRPVLLPADLRPAPPLP--------PLLQDLLPATRRRTTRTTTTGG--ADNGAQLHA 3803
```

```
                | :   :  ||:  |  :|        ||  ||  |  :   |     |   |  : |
Db       4509 AGNEKAVAVADMDWAAFIPAFTSVRPSPLFADLPEA--KAILRAAQDDGEDGDTASSLAD 4566

Qy       3804 RLAGQTHEQQHTTLLALVRSHIATVLGHTTPDTIPPDRAFRDLGFDSLTAVELRNRLSRT 3863
                 |       :|:  || ||| | :|||||:  : | | :||:::||||| || ||| |
Db       4567 SLRAVPDAEQNRILLKLVRGHASTVLGHSGAEGIGPRQAFQEVGFDSLAAVNLRNSLHAA 4626

Qy       3864 TGLRLPTTLAFDHPNPTTLTHHLHTQLQPQPDNAVAPVLAELDKLESALSALDKTDSASE 3923
                 |||||| || ||:| |  |  :|  :|  : |: :       | |    |:|:
Db       4627 TGLRLPATLIFDYPTPEALVGYLRVELLREADDGLD---GREDDLRRVLAAVPFARFKEA 4683

Qy       3924 RVTLRLKSLMLRWNAPQHPTAESADDDEKFTSATEAEIFKFID 3966
                   |  |  |      |      |:  |   :| :||:   :|
Db       4684 GVLDTLLGLADTGTEP----GTDAETTEAAPAADDAELIDALD 4722


RESULT 14
T30283
polyketide synthase - Streptomyces sp. (strain MA6548)
C;Species: Streptomyces sp.
A;Variety: strain MA6548
C;Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 17-Nov-2000
C;Accession: T30283
R;Motamedi, H.; Cai, S.J.; Shafiee, S.J.; Elliston, K.O.
Eur. J. Biochem. 244, 74-80, 1997
A;Title: Structural organization of a multifunctional polyketide synthase involved in th
A;Reference number: Z20806; MUID:97217427; PMID:9063448
A;Accession: T30283
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-6420 <MOT>
A;Cross-references: EMBL:Y10438; NID:e1014806; PID:e290681; PIDN:CAA71463.1
C;Genetics:
A;Note: fkbA
C;Superfamily: 3-oxoacyl-[acyl-carrier-protein] synthase I homology; acyl carrier protei
C;Keywords: carrier protein
F;51-433/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS1>
F;1930-2325/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS2>
F;3391-3462/Domain: acyl carrier protein homology <ACP1>
F;3505-3900/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS3>
F;3983-4254/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT1>
F;5307-5378/Domain: acyl carrier protein homology <ACP2>
F;5431-5831/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS4>
F;5932-6206/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT2>
F;6293-6364/Domain: acyl carrier protein homology <ACP3>

  Query Match             28.4%;  Score 5922;  DB 2;  Length 6420;
  Best Local Similarity   36.9%;  Pred. No. 4.3e-261;
  Matches 1641;  Conservative  408;  Mismatches 1310;  Indels 1090;  Gaps  140;

Qy         45 ELLDASDVFRESVRA-------CEAAFAPYVDWSVEQVLRDSPDAPGLDRVDVVQPTLFA 97
                |||||   | |  ||           |: |:          |  || | ||:
Db       1468 ELLDADGRFVEMGRAELRDPATIAPAYLPF----------DLLDA-GTDRI--------- 1507

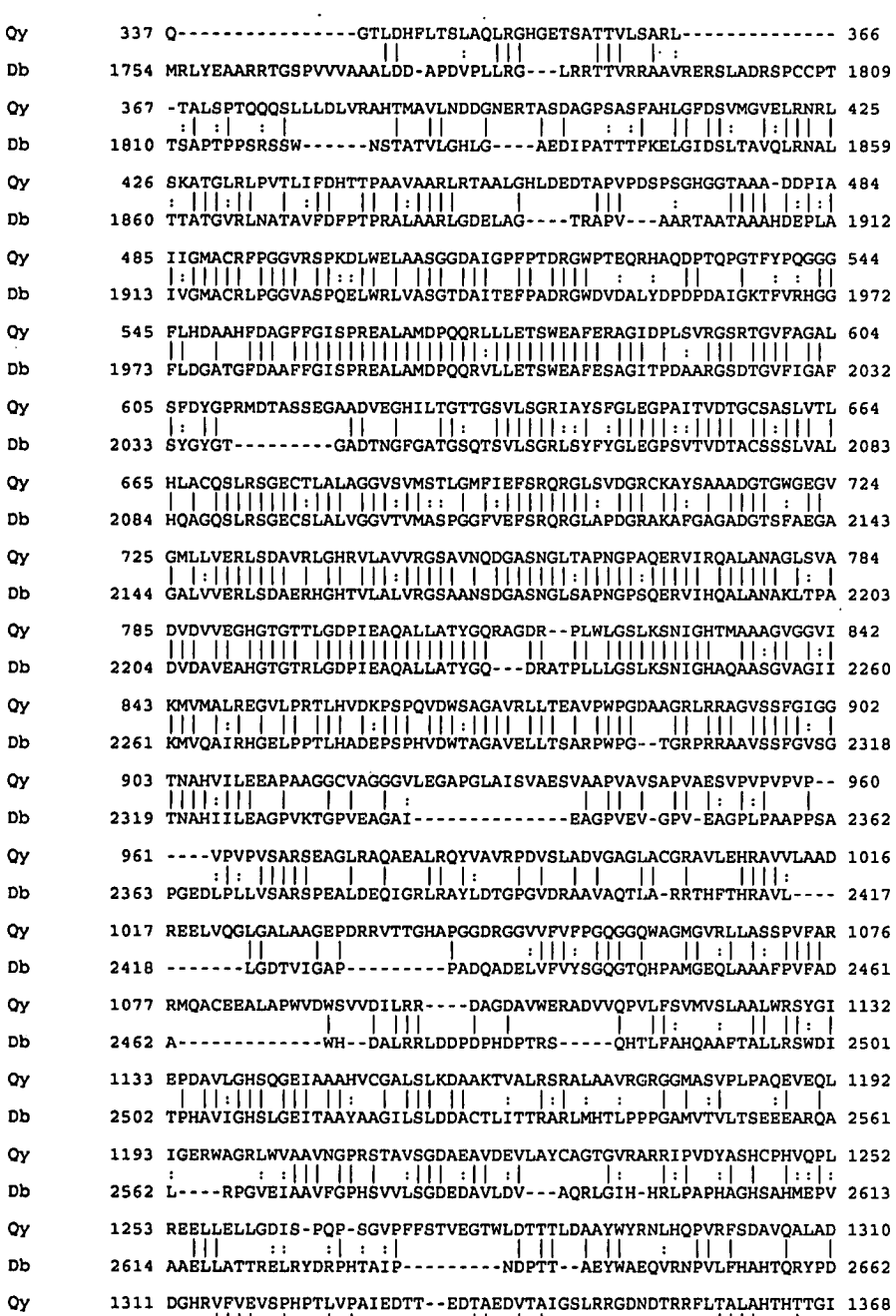
Qy         98 VMISLAALWRSQGVEPCAVLGHSLGEIAAAHVSGGLSLADAARVVTL----WS--QAQTT 151
                                   |  |||:        | | ||  :  |    |   ||:
Db       1508 --------------------GEILGEL--------LRLFDAGALTPLPVRAWDIRQARDA 1539

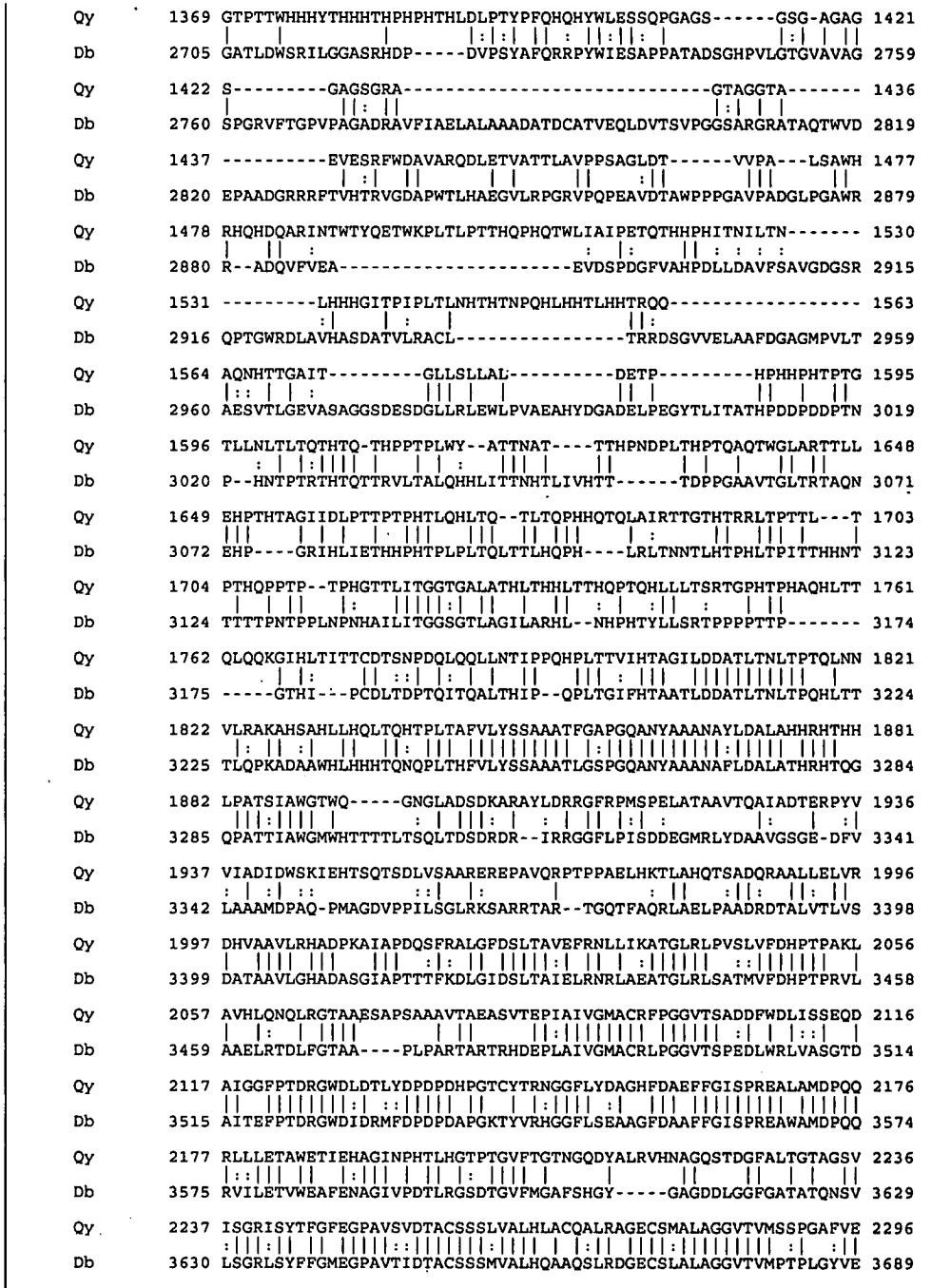
Qy        152 LA--------GTGALVSVAATPDELLPRIAPWTEDNPARLAVAAVNGPRSTVVSGAREAV 203
                |         |   |          |||                  ::   : |::|    :
Db       1540 LGVMSRARHIGKNVLT---------LPR--------------PLDPEGAIVITGGSGTL 1575

Qy        204 ADLVADLTAAQ---VRTRMIPVD-VPA-HSPLMYAIEERVVSGLLPITPRPSRIPFHS-- 256
                | ::|   :||   : :|  | :  |  | |    :  ::: : |  |  ||        |:
Db       1576 AGILARHVSAQHVYLLSRSAPPEGTPGIHLPCDVSDRQQLAAALRRI-DRPITAVVHTAG 1634

Qy        257 --------SVTGGRLDTR---ELDAAYWY-------RNMSSTVRFEPAARLLLQQGPKTF 298
                       |:| || ||    : | | |:       :::::  | :  || :|    |   :
Db       1635 ALDDGTLPSLTPGRFDTELRPKADGA-WHLHELTKDQDLAAFVLYSSAAGVLGNAGQGNY 1693

Qy        299 VEMS---------------PHPVLTMGLQELAPDLGDTTGTADTVIMG-------TLRRG 336
                :                |   :  || | |  |        ||   |         | |:|
Db       1694 AAANAFLDALAEQRRAAGLPALSIAWGLWEDASGLTAELSEADRQRMRRSGFRTITARQG 1753
```





```
Db       3690 FCRQRGLAPDGRAKAFAEGADGTSFSEGAGVLVVERLSDAERNGHTVLALVRSSAVNQDG 3749

Qy       2357 ASNGLTAPNGPSQQRVIRQALANAGLSAGDVDAVEAHGTGTTLGDPIEAQALLATYGQDR 2416
Db       3750 ASNGISAPNGPSQQRVIRQALDKAGLAPADVDVVEAHGTGTPLGDPIEAQAIIATYGQDR 3809

Qy       2417 AGEGPLWLGSVKSNVGHTQAAAGVAGVIKMVMALRHGLLPRTLHVDEPSPHVDWSAGAVQ 2476
Db       3810 --DTPLYLGSVKSNIGHTQTTAGLAGVIKMVMAMRHGLLPKTLHVDEPSSHVDWSAGAVE 3867

Qy       2477 LLTETVPWPGGEGRLRRAGVSSFGVSGTNAHVILEEAPADDVPGGPPAGEGDAGSDDEAA 2536
Db       3868 LLTEARPWPDSD-RPRRAGVSSLGISGTNAHVILEGVAESSV----------------RS 3910

Qy       2537 AGSPGVWPWLVSAKSQPALRAQAQALHAHLTDHPGLDLADVGYTLAHARAVFDHRATLIA 2596
Db       3911 VESSGLVPLPVSARTESSLALQVERLGEYV--RGARDLAAVADGLVRGRTVFGHRAVLL- 3967

Qy       2597 ADRDTFLQALQALAAGEPHPAVIHSSAPGGTGTGEAAGKTAFICSGQGTQRPGMAHGLYH 2656
Db       3968 --------------GESTVA----------GVAGAESRTVFVFPGQGSQWVGMGRELMG 4002

Qy       2657 THPVFAAALNDICTHLDPHLDHPLLPLLTQNDNDNEDAAALLQQTRYAQPALFAFQVALH 2716
Db       4003 CSEVFAARMRECAAVLEPYTGWDLLDVLGE--------AVVAERVEVLQPASWAVAVSL- 4053

Qy       2717 RLLTDGYHITPHYYAGHSLGEITAAHLAGILTLTDATTLITQRATLMQTMPP-----GTM 2771
Db       4054 AALWQAHGVSPDAVIGHSQGEIAAACVAGALSLEDAARIVALRS---QTIAAHLAGRGAM 4110

Qy       2772 TTLHTTPHHITHHLTAHE--NDLAIAAINTPTSLVISGTPHTVQHITTLCQQQGIKTKTL 2829
Db       4111 AS-------IALPATAVETVEGVWVAARNGPESTVVAGDPSAVERVLARYEAEGVRVRRI 4163

Qy       2830 PTNHAFHSPHTNPILNQLHQHTQTLTYHPPHTP--------LITANTPPDQLLTPHYWTQ 2881
Db       4164 AVDYASHTPHVEAIQEQLADVLGDITSSAPSVPWWSTVDGGWVTEPAGDD------YWYR 4217

Qy       2882 QARNTVDYATTTQTLHQHGVTTYIELGPDNTLTTLTHHNLPN-PPTTTLTLTHPHHHPQT 2940
Db       4218 NLRQPVAMDTAIGELDG---SLFIECSAHPVL-------LPALDQERTVASLRTDDGGWD 4267

Qy       2941 HLLTNLAKTTTTWHPHHYTHHDNQPHTHTHLDLPTYPFQHHHYWLESTQPGAGNVSAAGL 3000
Db       4268 RFLTALAQAWTQGADVDWTTL-IAPAPDRLLDLPTYPFDHKRYWIEAT--GAADLTALGL 4324

Qy       3001 DPTEHPLLGATLELATDGGALLAGRLSLRSHPWLADHAVGGTVLLSGATFLELALHAGTY 3060
Db       4325 TDTAHPMLAATAALPTNDGTVLTGRISLRTHPWLADHTIQDTVLLPGTAFLELVMRAGDE 4384

Qy       3061 VGCDRVDELTLHAPLVVPVDGGVSVQVGVAAADGEGRRLVSVYARGGSACGGGGASGGVW 3120
Db       4385 VGCDTIDELVIETPLTLPVTGAVDLTVTVEQADATGHRPVSVHAR--------PAGTDTW 4436

Qy       3121 TCHASGVLVEAAAGGVVVDGLAGV-----WPPRGAVAVDVDGVRDRLAGAGCVLGPVFSG 3175
Db       4437 TRHATGTL------GRTEDTAPDTNSFPQWPPTGPTALDVTEFYAQLAALGYRFVPTFRG 4490

Qy       3176 LRAVWRDGGDLLAEVCLPEEAWGDAAGFGLHPALLDGVVQPLSVLLPGGTGFGEGAGFGE 3235
Db       4491 LRAAWRAGDTVYAEVALPEDRAADADHFGVHPALLDAALQSGSLLMLESDG-------EQ 4543

Qy       3236 GVRVPAVWGGVSLHRAGVTGVRVRVSAVGRGGGREAVSVVVGDEAGVPVASVDRLELRPV 3295
Db       4544 GVQLPFSWHGVRFHATGATSLRVAIVP-----GPDGLRLRAADSENHPVATIDALVTRPP 4598

Qy       3296 DMGQLRAVSVSAGRRGSLYAVQWAEVGPV-PVCGQAWAWHEDVGESGGGPVPG------- 3347
Db       4599 ----------------------ADLAPSDPVLRVGWT---------AVPVPAEVRPSDA 4626

Qy       3348 -VVVLRCPDAGAGGGGGGGGGGVGEVVGGVLGVVQGWLGLERFAGSRLVVVTRGAVVAG 3406
```

```
Db       4627 DVLMLHSDDADP----------IQETRDLTTRVLDALLTTE---GTLIVQVTGG----- 4667

Qy       3407 PEDGPVDVVGASVWGLVRSAQAEHPDRFVLLDLDTDTGTDLDTGAGAGWGVDGGRVAAVV 3466
Db       4668 -------LAAMAAAGLVRTAQSEQPGRFFL--VETDPGEVL----------DGARCDMIA 4708

Qy       3467 ACGEPQLAVRGERLLAARLKR----------------------LESSGDVPAQR------ 3498
Db       4709 ALGEPRLRLRGNRFEAARLVRAAAALTPPDTGSWQLRQSATGSLDDVAIVPTGEPDRPLG 4768

Qy       3499 SGDTR---------------------------------------------------- 3503
Db       4769 AGEVRIAVRAAGLNFRDDTVALGVVADDRPLGSEAAGVVLETGPGVHDLAPGDRVLGMLA 4828

Qy       3504 ---------------------------------------------------------- 3503
Db       4829 GAFGPVAITDRRLLGRMPDGWTFPQAASVVTVFATTWYGLVDLAELRPGEKVLIHAAATG 4888

Qy       3504 --------ARRSDVPAQRSGGVPARRSVDV---------------------------- 3525
Db       4889 VGAAAVQIARHLDADVYATASTPKQHLLDLEPTHIADSRTTAFADTFPPADVVLNSLTGP 4948

Qy       3526 ---------------------------------------------------------- 3525
Db       4949 LLDASAQLLTPGGRFIEMGKTDIRHHVQQPFDLMVAGPDRLQQIITQLLDLFANGTLHPL 5008

Qy       3526 --------SGREVLPWLSG----------------GSVLVTGGTGVLGAAVARHLAGV 3559
Db       5009 PVHAWDVRQAREAFGWMSSGRHTGKLVLTVPRPLDPEGAVIITGGSGTLAGHLARHL--- 5065

Qy       3560 CGVRDLLLVSRRGP--DAPGAEGLRAELAALGAEVRIVACDVGERREVVRLLEGVPAGCP 3617
Db       5066 -DHPHTYLLSRTPPPDTTPGTH--------------LPCDVSDPTQLATALAQIPQ--P 5107

Qy       3618 LTGVVHAAGVLDDATIASLTPERLGTVFAAKVDAALLLDE--LTRGMELSAFVLFSSAAG 3675
Db       5108 LTGIFHTAATLDDALLHNLTPQRIDTVLQPKADAAWHLHHHTHTHATDLAAFVLYSAVAG 5167

Qy       3676 ILGSAGQGNYAAANAALDALAYRRRAAGLPGVSLAWGLWEEASGMTGHLAGTDHRRIIRS 3735
Db       5168 VMGSPGQGNYVAANAFLDALAEHRQAMGLPAQSLAWGMWADVSALTAKLTDADRGRIRRG 5227

Qy       3736 GLHPMSTPDALALFDAALALDRPVLLPADLRPAP------PLPPLLQDLLPATRRRTTRT 3789
Db       5228 GFPPLSAADGMRLFDSAA---RTRSRSSSRRPSDVTRLDGPVAPLLRGLV-AHRSGPVRA 5283

Qy       3790 TTTGGADNGAQLHARLAGQTHEQQHTTLLALVRSHIATVLGHTTPDTIPPDRAFRDLGFD 3849
Db       5284 VARNADDE--PLAARLAGHTAAEQRRVMQEVVLRQAAAVLAYGLGEQVAADRPFRDLGFD 5341

Qy       3850 SLTAVELRNRLSRTTGLRLPTTLAFDHPNPTTLTHHLHTQLQ---------PQPDNAVAP 3900
Db       5342 SLTAVDLRNRLAAETGLRLPTTVVFSHPTAEALATHLLELLDAPAALTTGAPLPAVTAAP 5401

Qy       3901 VLAELDKLE 3909
Db       5402 GTAARDQDE 5410
```

RESULT 15
S22012
erythronolide synthase (EC 2.3.1.94) III - Saccharopolyspora erythraea
N;Alternate names: 6-deoxyerythronolide B synthase III; protein A
C;Species: Saccharopolyspora erythraea
C;Date: 07-Apr-1994 #sequence_revision 07-Apr-1994 #text_change 26-May-2000
C;Accession: S22012; S34784; S23206
R;Bevitt, D.J.
submitted to the EMBL Data Library, September 1991
A;Reference number: S22011
A;Accession: S22012
A;Molecule type: DNA
A;Residues: 1-3172 <BEV1>
A;Cross-references: EMBL:X62569; NID:g46977; PIDN:CAA44449.1; PID:g46979

```
A;Experimental source: strain NRRL 2338
R;Bevitt, D.J.; Cortes, J.; Haydock, S.F.; Leadlay, P.F.
Eur. J. Biochem. 204, 39-49, 1992
A;Title: 6-Deoxyerythronolide-B synthase 2 from Saccharopolyspora erythraea. Cloning of
A;Reference number: S23070; MUID:92155230; PMID:1740151
A;Accession: S34784
A;Molecule type: DNA
A;Residues: 1-715,717-728,'C',729-886,888-894,'QR',895-910,'E',912-1276,'RR',1279-1284,'
EV2>
A;Cross-references: EMBL:X62569
R;Caffrey, P.; Bevitt, D.J.; Staunton, J.; Leadlay, P.F.
FEBS Lett. 304, 225-228, 1992
A;Title: Identification of DEBS 1, DEBS 2 and DEBS 3, the multienzyme polypeptides of th
A;Reference number: S23103; MUID:92316235; PMID:1618327
A;Accession: S23206
A;Molecule type: protein
A;Residues: 2-11,'XXX',15 <CAF>
A;Experimental source: strain CA340
C;Genetics:
A;Gene: eryA
C;Function:
A;Description: catalyzes the construction of a polyketide chain, which is then cyclised
A;Pathway: erythromycin biosynthesis
C;Superfamily: [acyl-carrier-protein] S-malonyltransferase homology; 3-oxoacyl-[acyl-car
ogy; short-chain alcohol dehydrogenase homology
C;Keywords: acyltransferase; antibiotic biosynthesis; carrier protein; coenzyme A; multi
F;60-450/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS1>
F;555-835/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT1>
F;1117-1297/Domain: short-chain alcohol dehydrogenase homology <SAD1>
F;1392-1463/Domain: acyl carrier protein homology <ACP1>
F;1511-1914/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS2>
F;2022-2300/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT2>
F;2557-2734/Domain: short-chain alcohol dehydrogenase homology <SAD2>
F;2819-2890/Domain: acyl carrier protein homology <ACP2>
F;2959-3153/Domain: oleoyl-[acyl-carrier-protein] hydrolase homology <ACPH>

  Query Match            25.9%;  Score 5392;  DB 2;  Length 3172;
  Best Local Similarity  38.5%;  Pred. No. 2.4e-237;
  Matches 1343;  Conservative  357;  Mismatches 1082;  Indels  706;  Gaps   77;

Qy        479 ADDPIAIIGMACRFPGGVRSPKDLWELAASGGDAIGPFPTDRGWPTEQRHAQDPTQPGTF 538
Db         37 AGEPIAIVGMACRFPGDVDSPESFWEFVSGGGDAIAEAPADRGW------EPDPD----- 85

Qy        539 YPQGGGFLHDAAHFDAGFFGISPREALAMDPQQRLLLETSWEAFERAGIDPLSVRGSRTG 598
Db         86 -ARLGGMLAAAGDFDAGFFGISPREALAMDPQQRIMLEISWEALERAGHDPVSLRGSATG 144

Qy        599 VFAGALSFDYGPRMDTASSEGAADVEGHILTGTTGSVLSGRIAYSFGLEGPAITVDTGCS 658
Db        145 VFTGVGTVDYGPRPDEAPDE----VLGYVGTGTASSVASGRVAYCLGLEGPAMTVDTACS 200

Qy        659 ASLVTLHLACQSLRSGECTLALAGGVSVMSTLGMFIEFSRQRGLSVDGRCKAYSAAADGT 718
Db        201 SGLTALHLAMESLRRDECGLALAGGVTVMSSPGAFTEFRSQGGLAADGRCKPFSKAADGF 260

Qy        719 GWGEGVGMLLVERLSDAVRLGHRVLAVVRGSAVNQDGASNGLTAPNGPAQERVIRQALAN 778
Db        261 GLAEGAGVLVLQRLSAARREGRPVLAVLAGSAVNQDGASNGLTAPSGPAQQRVIRRALEN 320

Qy        779 AGLSVADVDVVEGHGTGTTLGDPIEAQALLATYG-QRAGDRPLWLGSLKSNIGHTMAAAG 837
Db        321 AGVRAGDVDYVEAHGTGTRLGDPIEVHALLSTYGAERDPDDPLWIGSVKSNIGHTQAAAG 380

Qy        838 VGGVIKMVMALREGVLPRTLHVDKPSPQVDWSAGAVRLLTEAVPWPGDAAGRLRRAGVSS 897
Db        381 VAGVMKAVLALRHGEMPRTLHFDEPSPQIEWDLGAVSVVSQARSWP--AGERPRRAGVSS 438

Qy        898 PGIGGTNAHVILEEAPAAGGCVAGGGVLEGAPGLAISVAESVAAPVAVSAPVAESVPVPV 957
Db        439 PGISGTNAHVIVEEAPEAD--------------------------------EPEPAPD 464

Qy        958 PVPVPVPVSARSEAGLRAQAEALRQYVAVRPDVSLADVGAGLACGRAVLEHRAVVLAADR 1017

Db        465 SGPVPLVLSGRDEQAMRAQAGRLADHLAPEPRNSLRDTGFTLATRASAMEHRAVVV-GDR 523

Qy       1018 EELVQGLGALAAGEPDRRV----TTGHAPGGDRGGVVFVFPGQGGQWAGMGVRLLASSPV 1073
Db        524 DEALAGLRAVA----DRRIADRTATGQGPNSPR-RVAMVFPGQGAQWQGMARDLLRESQV 578

Qy       1074 FARRMQACEEALAPWVDWSVVDILRRDAGDAVWERADVVQPVLFSVMVSLAALWRSYGIE 1133
Db        579 FADSIRDCERALAPHVDWSLTDLL---SGARPLDRVDVVQPALFAVMVSLAALWRSHGVE 635

Qy       1134 PDAVLGHSQGEIAAAHVCGALSLKDAAKTVALRSRALAAVRGRGGMASVPLPAQEVEQLI 1193
Db        636 PAAVVGHSQGEIAAAHVAGALTLEDAAKLVAVRSRVLRRLGGQGGMASFGLGTEQAAERI 695

Qy       1194 GERWAGRLWVAAVNGPRS-TAVSGDAEAVDEVLAYCAGTGVRARRIPVDYASHCPHVQPL 1252
Db        696 G-RFAGALSIASVNGPRSVVVVAGESGPLDELIAECEAEAHKARRIPVDYASHSPQVESL 754

Qy       1253 REELLELLGDISPQPSGVPFFSTVEGTWLDTTTLDAAYWYRNLHQPVRFSDAVQALADDG 1312
Db        755 REELLTELAGISPVSADVALYSTTTGQPIDTATMDTAYWYANLREQVRFQDATRQLAEAG 814

Qy       1313 HRVFVEVSPHPTLVPAIEDTTEDTAEDVTA----IGSLRRGDNDTRRFLTALAHTHTTGI 1368
Db        815 FDAFVEVSPHPVLTVGIE-ATLDSALPADAGACVVGTLRRDRGGLADFHTALGEAYAQGV 873

Qy       1369 GTPTTWHHHYTHHHTHPHPHTHLDLPTYPFQHQHYWLESSQPGAGSGSGAGAGSGAGSGR 1428
Db        874 --EVDWSPAFA--DARP-----VELPVYPF--QRYWLPI---------------------- 901

Qy       1429 AGTAGGTAEVESRFWDAVARQDLETVATTLAVPPSAGLDTVVPALSAWHRHQHDQARINT 1488
Db        902 --PTGGRARDED-----------------------------------------------DD 913

Qy       1489 WTYQETWKPLTLPTTHQPHQTWLIAIPETQTHHPHITNILTNLHHHGITPIPLTLNHTHT 1548
Db        914 WRYQVVWREAEWESASLAGRVLLVTGPGVPSELSDA--IRSGLEQSGATVLTCDV----- 966

Qy       1549 NPQHLHHTLHHTRQQAQNHTTGAITGLLSLLALDETPHPHHPHTPTGTLLNLTLTQTHTQ 1608
Db        967 ----------ESRSTIGTALEAADTDALSTVGVAAVP---HGEAVDPSLDALALVQALGA 1013

Qy       1609 THPPTPLWYATTNATTTHPNDPLTHPTQAQTWGLARTTLLEHPTHTAGIIDLPTTPTPHT 1668
Db       1014 AGVEAPLWVLTRNAVQVADGE-LVDPAQAMVGGLGRVVGIEQPGRWGGLVDLVDADAA-S 1071

Qy       1669 LQHLTQTLTQPHHQTQLAIRTTGTHTRRLTPTTLTPTHQPPTPTPHGTTLITGGTGALAT 1728
Db       1072 IRSLAAVLADPRGEEQVAIRADGIKVARLVPAPARARTHPLEPLA-GTVLVTGGTGGIGA 1130

Qy       1729 HLTHHLTTHQPTQHLLLTSRTGPHTPHAQHLTTQLQQKGIHLTITTCDTSNPDQLQQLL- 1787
Db       1131 HLARWL-ARSGAEHLVLLGRRGADAPGASELREELTALGTGVTIAACDVADRARLEAVLA 1189

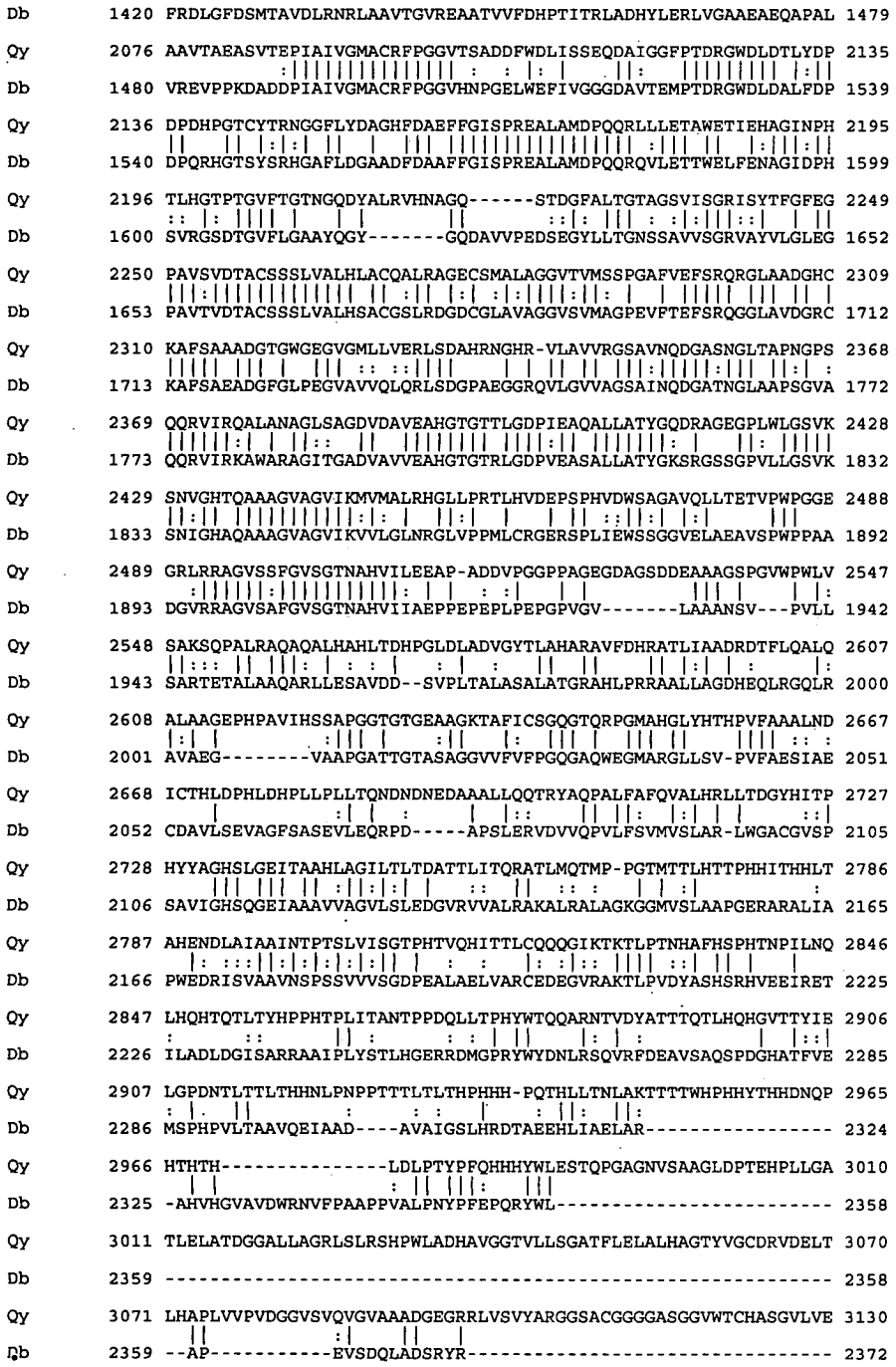
Qy       1788 -NTIPPQHPLTTVIHTAGILDDATLTNLTPTQLNNVLRAKAHSAHLLHQLTQHTPLTAFV 1846
Db       1190 AEAAAEGRTVSAVMHAAGVSTSTPLDDLTEAEFTEIADVKVRGTVNLDELC--PDLDAFV 1247

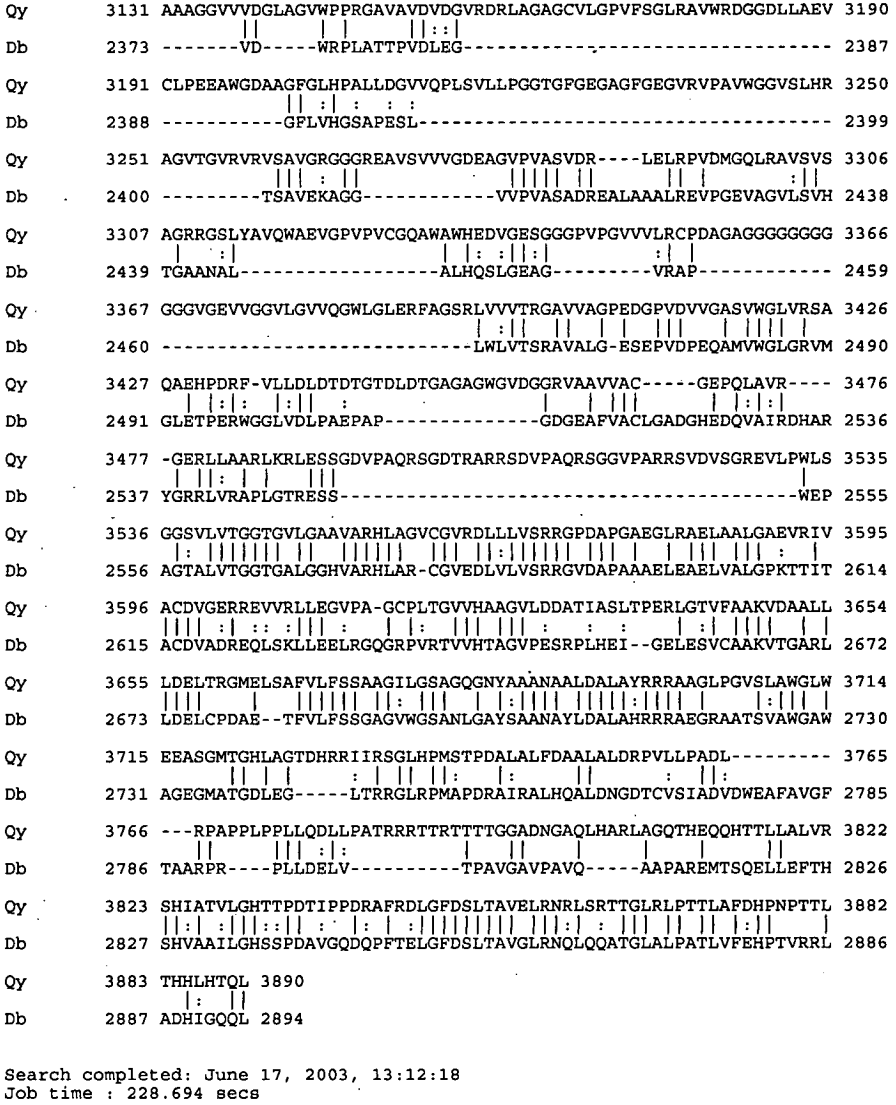
Qy       1847 LYSSAAATFGAPGQANYAAANAYLDALAHHRHTHHLPATSIAWGTWQGNGLADSDKARAY 1906
Db       1248 LFSSNAGVWGSPGLASYAAANAFLDGFARAARSEGAPVTSIAWGLWAGQNMA-GDEGGEY 1306

Qy       1907 LDRRGFRPMSPELATAAVTQAIADTERPYVVIADIDWSKIEHTSQTSDLVSAARE----- 1961
Db       1307 LRSQGLRAMDPDRAVEEL-HITLDHGQTSVSVVDMDRRRF------VELFTAARHRPLFD 1359

Qy       1962 -----REPAVQRPTPPAELHKTLAHQTSADQRAALLELVRDHVAAVLRHADPKAIAPDQS 2016
Db       1360 EIAGARAEARQSEEGPALAQRLAALLCDGRREREHLAHLIRAEVAAVLGHGDDAAIDRDRA 1419

Qy       2017 FRALGFDSLTAVEFRNLLIKATGLRLPVSLVFDHPTPAKLAVHLQNQLRGTA-AESAPSA 2075
```





Search completed: June 17, 2003, 13:12:18
Job time : 228.694 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 17, 2003, 12:52:22 ; Search time 43.3159 Seconds
(without alignments)
3803.313 Million cell updates/sec

Title: US-09-914-286-3
Perfect score: 20829
Sequence: 1 VQRMDGGEEPRPAAGEVLGV.....FTSATEAEIFKFIDNDLGLS 3972

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	6387	30.7	3567	1	ERY2_SACER	Q03132 saccharopol
2	6360	30.5	3491	1	ERY1_SACER	Q03131 saccharopol
3	6108.5	29.3	3519	1	OL56_STRAT	Q07017 streptomyce
4	5392	25.9	3172	1	ERY3_SACER	Q03133 saccharopol
5	2050.5	9.8	1538	1	PPSB_MYCTU	Q10978 mycobacteri
6	2042	9.8	4427	1	PKSL_BACSU	Q05470 bacillus su
7	2017	9.7	1876	1	PPSA_MYCTU	Q10977 mycobacteri
8	1832.5	8.8	2110	1	MCAS_MYCBO	Q02251 mycobacteri
9	1809	8.7	4447	1	PKSK_BACSU	P40803 bacillus su
10	1652	7.9	4273	1	PKSM_BACSU	P40872 bacillus su
11	1638.5	7.9	1774	1	MSAS_PENPA	P22367 penicillium
12	1117.5	5.4	2109	1	PKS1_ASPPA	Q12053 aspergillus
13	1064.5	5.1	1986	1	WA_EMENI	Q03149 emericella
14	1019.5	4.9	2181	1	STCA_EMENI	Q12397 emericella
15	1010	4.8	2511	1	FAS_CHICK	P12276 gallus gall
16	964.5	4.6	2505	1	FAS_RAT	P12785 rattus norv
17	895	4.3	2504	1	FAS_HUMAN	P49327 homo sapien
18	628	3.0	352	1	FAS_ANSAN	P36189 anser anser
19	370.5	1.8	401	1	NOE4_RHIME	P06231 rhizobium m
20	368	1.8	402	1	NODE_RHIME	P06230 rhizobium m
21	365	1.8	426	1	KAS1_STRGA	P16538 streptomyce
22	358.5	1.7	401	1	NODE_RHILT	P04684 rhizobium l
23	354	1.7	402	1	NODE_RHIS3	P72331 rhizobium s
24	352.5	1.7	424	1	KASA_STRCO	Q02059 streptomyce
25	352	1.7	403	1	NODE_RHILV	P04683 rhizobium l
26	342	1.6	421	1	FABF_RHIME	P56902 rhizobium m
27	335	1.6	412	1	FABF_ECOLI	P39435 escherichia
28	326	1.6	420	1	KAS1_STRCM	P41175 streptomyce
29	322.5	1.5	422	1	KAS1_STRCN	Q02578 streptomyce
30	319	1.5	417	1	FAB2_MYCTU	Q10525 mycobacteri
31	317.5	1.5	442	1	CEM1_YEAST	P39525 saccharomyc
32	312.5	1.5	422	1	KAS1_STRHA	Q05356 streptomyce
33	311	1.5	462	1	FABB_HORVU	P23902 hordeum vul
34	308.5	1.5	413	1	FABF_VIBCH	Q9kqh9 vibrio chol
35	308	1.5	421	1	KAS1_STRVN	P16540 streptomyce
36	307	1.5	414	1	FABF_VIBHA	P55338 vibrio harv
37	298	1.4	416	1	FAB1_MYCTU	Q10524 mycobacteri
38	296.5	1.4	473	1	FABB_ARATH	P52410 arabidopsis
39	296	1.4	416	1	PKSF_BACSU	P40804 bacillus su
40	295.5	1.4	420	1	FAB2_MYCLE	O69473 mycobacteri
41	291	1.4	416	1	FAB1_MYCLE	Q9cbs7 mycobacteri
42	290.5	1.4	423	1	KAS1_STRCO	P23155 streptomyce
43	285	1.4	425	1	KAS1_STRRM	P43678 streptomyce
44	281.5	1.4	406	1	FABB_ECOLI	P14926 escherichia
45	280.5	1.3	416	1	FABF_SYNY3	P73283 synechocyst

ALIGNMENTS

RESULT 1
ERY2_SACER
ID ERY2_SACER STANDARD; PRT; 3567 AA.
AC Q03132; Q54096;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Erythronolide synthase, modules 3 and 4 (EC 2.3.1.94) (ORF 2) (6-
DE deoxyerythronolide B synthase II) (DEBS 2).
GN ERYA.
OS Saccharopolyspora erythraea (Streptomyces erythraeus).
OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC Actinomycetales; Pseudonocardineae; Pseudonocardiaceae;
OC Saccharopolyspora.
OX NCBI_TaxID=1836;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91220065; PubMed=2024119;
RA Donadio S., Staver M.J., McAlpine J.B., Swanson S.J., Katz L.;
RT "Modular organization of genes required for complex polyketide
RT biosynthesis.";
RL Science 252:675-679(1991).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=NRRL 2338;
RX MEDLINE=92155230; PubMed=1740151;
RA Bevitt D.J., Cortes J., Haydock S.F., Leadlay P.F.;
RT "6-Deoxyerythronolide-B synthase 2 from Saccharopolyspora erythraea.
RT Cloning of the structural gene, sequence analysis and inferred domain
RT structure of the multifunctional enzyme.";
RL Eur. J. Biochem. 204:39-49(1992).
CC -!- CATALYTIC ACTIVITY: 6 malonyl-CoA + propionyl-CoA = 7 CoA + 6-
CC deoxyerythronolide B.
CC -!- COFACTOR: NADP; CONTAINS 2 COVALENTLY BOUND PHOSPHOPANTETHEINES.
CC -!- PATHWAY: COMPLEX POLYKETIDE FORMATION IN ERYTHROMYCIN
CC BIOSYNTHESIS.
CC -!- MISCELLANEOUS: IN EACH ORF OF ERYA TWO MODULES ARE PRESENT EACH
CC ENCODING FOR A FUNCTIONAL SYNTHASE SUBUNIT. THUS ERYA SHOWING 3
CC ORFS CODES FOR 6 SYNTHASE SUBUNITS. IT IS SUPPOSED THAT EACH
CC SYNTHASE PARTICIPATES IN ONE OF THE SIX FAS-LIKE ELONGATION STEPS
CC REQUIRED FOR FORMATION OF THE POLYKETIDE. MODULE 1, 2, 3, 4, 5,
CC AND 6 PARTICIPATING IN BIOSYNTHESIS STEPS 1, 2, 3, 4, 5, AND 6,
CC RESPECTIVELY.
CC -!- MISCELLANEOUS: BIOSYNTHESIS OF POLYKETIDES; ACYLTRANSFERASE (AT),
CC BETA-KETOACYL CARRIER PROTEIN SYNTHASE (KS), AND ACYL CARRIER
CC PROTEIN (ACP) FOR CHAIN ELONGATION. BETA-KETOREDUCTASE (KR),
CC DEHYDRATASE (DH), AND ENOYL REDUCTASE (ER) FOR PROCESSING OF THE
CC BETA CARBON, AND THIOESTERASE (TE) FOR RELEASE AND LACTONIZATION
CC OF THE FULL-LENGTH CHAIN.
CC -!- SIMILARITY: TO FATTY ACID SYNTHASE (FAS).
CC -!- SIMILARITY: CONTAINS 2 ACYL CARRIER DOMAINS.
CC ----------
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way

```
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; M63677; AAA26494.1; -.
DR   EMBL; X62569; CAA44448.1; -.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR002085; Adh_zn_family.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR003880; Ppantne_attach.
DR   Pfam; PF00107; adh_zinc; 1.
DR   Pfam; PF00109; ketoacyl-synt; 2.
DR   Pfam; PF00550; pp-binding; 2.
DR   Pfam; PF00698; Acyl_transf; 2.
DR   Pfam; PF02801; ketoacyl-synt_C; 2.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; 2.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 2.
DR   PROSITE; PS50075; ACP_DOMAIN; 2.
KW   Transferase; Acyltransferase; Antibiotic biosynthesis; NADP; Repeat;
KW   Phosphopantetheine; Multifunctional enzyme.
FT   DOMAIN        1   1484       MODULE 3.
FT   DOMAIN     1485   3567       MODULE 4.
FT   DOMAIN       27    488       BETA-KETOACYL SYNTHASE 1.
FT   DOMAIN      559    884       ACYLTRANSFERASE (AT) 1.
FT   DOMAIN     1130   1301       BETA-KETOACYL REDUCTASE 1 (POSSIBLY
FT                                NON-FUNCTIONAL).
FT   DOMAIN     1397   1467       ACYL CARRIER (ACP) 1.
FT   DOMAIN     1485   1943       BETA-KETOACYL SYNTHASE 2.
FT   DOMAIN     2013   2336       ACYLTRANSFERASE (AT) 2.
FT   DOMAIN     2383   3066       DEHYDRATASE/ENOYLREDUCTASE (DH/ER).
FT   DOMAIN     3139   3322       BETA-KETOACYL REDUCTASE 2.
FT   DOMAIN     3415   3485       ACYL CARRIER (ACP) 2.
FT   ACT_SITE    202    202       THIOESTER BOND.
FT   ACT_SITE    651    651       ACYL-ENZYME INTERMEDIATE.
FT   BINDING    1430   1430       PHOSPHOPANTETHEINE (BY SIMILARITY).
FT   ACT_SITE   1661   1661       THIOESTER BOND.
FT   ACT_SITE   2115   2115       ACYL-ENZYME INTERMEDIATE.
FT   NP_BIND    2961   2978       NADP (ER).
FT   NP_BIND    3142   3157       NADP (KR).
FT   BINDING    3448   3448       PHOSPHOPANTETHEINE (BY SIMILARITY).
FT   CONFLICT    438    438       R -> A (IN REF. 2).
FT   CONFLICT    480    480       T -> S (IN REF. 2).
FT   CONFLICT   1241   1241       L -> F (IN REF. 2).
FT   CONFLICT   2664   2664       G -> V (IN REF. 2).
SQ   SEQUENCE   3567 AA;  374413 MW;  EE6284F4738AA0C0 CRC64;

  Query Match             30.7%;  Score 6387;  DB 1;  Length 3567;
  Best Local Similarity   40.1%;  Pred. No. 6.8e-283;
  Matches 1531;  Conservative  435;  Mismatches 1242;  Indels  606;  Gaps   69;

Qy        481 DPIAIIGMACRFPGGVRSPKDLWELAASGGDAIGPFPTDRGWPTEQRHAQDPTQPGTFYP 540
Db         31 DPIAIVSMACRLPGGVNTPQRLWELLREGGETLSGFPTDRGWDLARLHHPDPDNPGTSYV 90

Qy        541 QGGGFLHDAAHFDAGFFGISPREALAMDPQQRLLLETSWEAFERAGIDPLSVRGSRTGVF 600
Db         91 DKGGFLDDAAGFDAEFFGVSPREAAAMDPQQRLLLETSWELVENAGIDPHSLRGTATGVF 150

Qy        601 AGALSFDYGPRMDTASSEGAADVEGHILTGTTGSVLSGRIAYSFGLEGPAITVDTGCSAS 660
Db        151 LGVAKFGYGE--DTAAAE---DVEGYSVTGVAPAVASGRISYTMGLEGPSISVDTACSSS 205

Qy        661 LVTLHLACQSLRSGECTLALAGGVSVMSTLGMFIEFSRQRGLSVDGRCKAYSAAADGTGW 720
Db        206 LVALHLAVESLRKGESSMAVVGGAAVMATPGVFVDFSRQRALAADGRSKAFGAGADGFGF 265

Qy        721 GEGVGMLLVERLSDAVRLGHRVLAVVRGSAVNQDGASNGLTAPNGPAQERVIRQALANAG 780
Db        266 SEGVTLVLLERLSEARRNGHEVLAVVRGSALNQDGASNGLSAPSGPAQRRVIRQALESCG 325

Qy        781 LSVADVDVVEGHGTGTTLGDPIEAQALLATYGQ-RAGDRPLWLGSLKSNIGHTMAAAGVG 839
Db        326 LEPGDVDAVEAHGTGTALGDPIEANALLDTYGRDRDADRPLWLGSVKSNIGHTQAAAGVT 385

Qy        840 GVIKMVMALREGVLPRTLHVDKPSPQVDWSAGAVRLLTEAVPWPGDAAGRLRRAGVSSFG 899
Db        386 GLLKVVLALRNGELPATLHVEEPTPHVDWSSGGVALLAGNQPW--RRGERTRRARVSAFG 443

Qy        900 IGGTNAHVILEEAPAAGGCVAGGGVLEGAPGLAISVAESVAAPVAVSAPVAESVPVPVPV 959
Db        444 ISGTNAHVIVEEAPER----------------EHRETTAHDGR--------------- 470

Qy        960 PVPVPVSARSEAGLRAQAEALRQYVAVRPDVSLADVGAGLACGRAVLEHRAVVLAADREE 1019
Db        471 PVPLVVSARTTAALRAQAAQIAELLE-RPDADLAGVGLGLATTRARHEHRAAVVASTREE 529

Qy       1020 LVQGLGALAAGEPDRRVTTGHAPGGDRGGVVFVFPGQGGQWAGMGVRLLASSPVFARRMQ 1079
Db        530 AVRGLREIAAGAATADAVVEGVTEVDGRNVVFLFPGQGSQWAGMGAELLSSSPVFAGKIR 589

Qy       1080 ACEEALAPWVDWSVVDILRRDAGDAVWERADVVQPVLFSVMVSLAALWRSYGIEPDAVLG 1139
Db        590 ACDESMAPMQDWKVSDVLRQAPGAPGLDRVDVVQPVLFAVMVSLAELWRSYGVEPAAVVG 649

Qy       1140 HSQGEIAAAHVCGALSLKDAAKTVALRSRALAAVRGRGGMASVPLPAQEVEQLIGERWAG 1199
Db        650 HSQGEIAAAHVAGALTLEDAAKLVVGRSRLMRSLSGEGGMAAVALGEAAVRERL-RPWQD 708

Qy       1200 RLWVAAVNGPRSTAVSGDAEAVDEVLAYCAGTGVRARRIPVDYASHCPHVQPLREELLEL 1259
Db        709 RLSVAAVNGPRSVVVSGEPGALRAFSEDCAAEGIRVRDIDVDYASHSPQIERVREELLET 768

Qy       1260 LGDISPQPSGVPFFSTVEGTWLDTTTLDAAYWYRNLHQPVRFSDAVQALADDGHRVFVEV 1319
Db        769 TGDIAPRPARVTFHSTVESRSMDGTELDARYWYRNLRETVRFADAVTRLAESGYDAFIEV 828

Qy       1320 SPHPTLVPAIEDTTE--DTAEDVTAIGSLRRGDNDTRRFLTALAHTHTTGI--------- 1368
Db        829 SPHPVVVQAVEEAVEEADGAEDAVVVGSLHRDGGDLSAFLRSMATAHVSGVDIRWDVALP 888

Qy       1369 -GTPTTWHHHYTHHHTHPHPHTHLDLPTYPFQHQHYWLESSQPGAGSGSGAGAGSGAGSG 1427
Db        889 GAAP------------------FALPTYPFQRKRYWLQPAAPAAAS------------ 916

Qy       1428 RAGTAGGTAEVESRFWDAVARQDLETVATTLAVPPSAGLDTVVPALSAWHRHQHDQARIN 1487
Db        917 ---------------DELA---------------------------------------- 920

Qy       1488 TWTYQETWKPLTLPTTHQPHQTWLIAIPETQTHHPHITNILTN-LHHHGITPIPLTLNHT 1546
Db        921 ---YRVSWTPIEKPESGNLDGDWLVVTPLIS---PEWTEMLCEAINANGGRALRCEVDTS 974

Qy       1547 HTNPQHLHHTLHHTRQQAQNHTTGAITGLLSLLALDETPHPHHPHTPTGTLLNLTLTQTH 1606
Db        975 ASRTE---------MAQAVAQAGTGFRGVLSLLSSDES--ACRPGVPAGAVGLLTLVQAL 1023

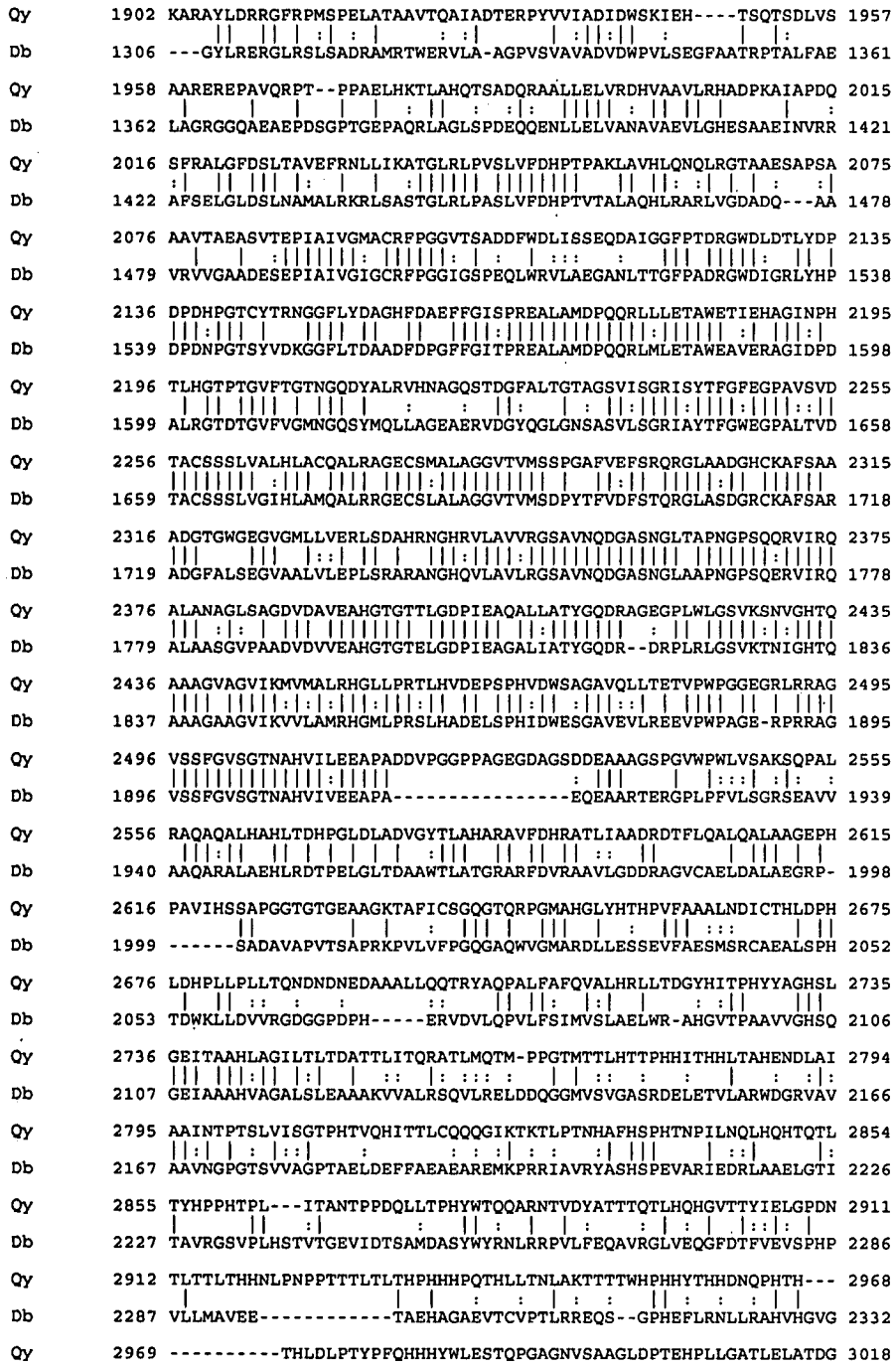
Qy       1607 TQTHPPTPLWYATTNATTTHPNDPLTHPTQAQTWGLARTTLLEHPTHTAGIIDLPTTPTP 1666
Db       1024 GDAGVDAPVWCLTQGAVRTPADDDLARPAQTTAHGFAQVAGLELPGRWGGVVDLPESVDD 1083

Qy       1667 HTLQHLTQTL-----TQPHHQTQLAIRTTGTHTRRLTPTTLTPTHQPPTPTPHGTTLITG 1721
Db       1084 AALRLLVAVLRGGGRAEDH----LAVRDGRLHGRRVVRASL-PQSGSRSWTPHGTVLVTG 1138

Qy       1722 GTGALATHLTHHLTTHQPTQHLLLTSRTGPHTPHAQHLTTQLQQKGIHLTITTCDTSNPD 1781
Db       1139 AASPVGDQLVRWLADRGAERLVLAGACPG------DDLLAAVEEAGASAVVCAQDAA--- 1189

Qy       1782 QLQQLLNTIPPQHPLTTVIHTAGILDDATLTNLTPTQLNNVLRAKAHSAHLLHQLTQHTP 1841
Db       1190 ALREALG----DEPVTALVHAGTLTNFGSISEVAPEEFAETIAAKTALLAVLDEVLGDRA 1245

Qy       1842 LTAFVLYSSAAATFGAPGQANYAAANAYLDALAHHRHTHHLPATSIAWGTWQGNGLADSD 1901
Db       1246 VEREVYCSSVAGIWGGAGMAAYAAGSAYLDALAEHHRARGRSCTSVAWTPWALPGGAVDD 1305
```


Qy 1902 KARAYLDRRGFRPMSPELATAAVTQAIADTERPYVVIADIDWSKIEH----TSQTSDLVS 1957
Db 1306 ---GYLRERGLRSLSADRAMRTWERVLA-AGPVSVAVADVDWPVLSEGFAATRPTALFAE 1361
Qy 1958 AARAREPAVQRPT--PPAELHKTLAHQTSADQRAALLELVRDHVAAVLRHADPKAIAPDQ 2015
Db 1362 LAGRGGQAEAEPDSGPTGEPAQRLAGLSPDEQQENLLELVANAVAEVLGHESAAEINVRR 1421
Qy 2016 SFRALGFDSLTAVEFRNLLIKATGLRLPVSLVFDHPTPAKLAVHLQNQLRGTAAESAPSA 2075
Db 1422 AFSELGLDSLNAMALRKRLSASTGLRLPASLVFDHPTVTALAQHLRARLVGDADQ---AA 1478
Qy 2076 AAVTAEASVTEPIAIVGMACRFPGGVTSADDFWDLISSEQDAIGGFPTDRGWDLDTLYDP 2135
Db 1479 VRVVGAADESEPIAIVGIGCRFPGGIGSPEQLWRVLAEGANLTTGFPADRGWDIGRLYHP 1538
Qy 2136 DPDHPGTCYTRNGGFLYDAGHFDAEFFGISPREALAMDPQQRLLLETAWETIEHAGINPH 2195
Db 1539 DPDNPGTSYVDKGGFLTDAADFDPGFFGITPREALAMDPQQRLMLETAWEAVERAGIDPD 1598
Qy 2196 TLHGTPTGVFTGTNGQDYALRVHNAGQSTDGFALTGTAGSVISGRISYTFGFEGPAVSVD 2255
Db 1599 ALRGTDTGVFVGMNGQSYMQLLAGEAERVDGYQGLGNSASVLSGRIAYTFGWEGPALTVD 1658
Qy 2256 TACSSSLVALHLACQALRAGECSMALAGGVTVMSSPGAFVEFSRQRGLAADGHCKAFSAA 2315
Db 1659 TACSSSLVGIHLAMQALRRGECSLALAGGVTVMSDPYTFVDFSTQRGLASDGRCKAFSAR 1718
Qy 2316 ADGTGWGEGVGMLLVERLSDAHRNGHRVLAVVRGSAVNQDGASNGLTAPNGPSQQRVIRQ 2375
Db 1719 ADGFALSEGVAALVLEPLSRARANGHQVLAVLRGSAVNQDGASNGLAAPNGPSQERVIRQ 1778
Qy 2376 ALANAGLSAGDVDAVEAHGTGTTLGDPIEAQALLATYGQDRAGEGPLWLGSVKSNVGHTQ 2435
Db 1779 ALAASGVPAADVDVVEAHGTGTELGDPIEAGALIATYGQDR--DRPLRLGSVKTNIGHTQ 1836
Qy 2436 AAAGVAGVIKMVMALRHGLLPRTLHVDEPSPHVDWSAGAVQLLTETVPWPGGEGRLRRAG 2495
Db 1837 AAAGAAGVIKVVLAMRHGMLPRSLHADELSPHIDWESGAVEVLREEVPWPAGE-RPRRAG 1895
Qy 2496 VSSFGVSGTNAHVILEEAPADDVPGGPPAGEGDAGSDDEAAAGSPGVWPWLVSAKSQPAL 2555
Db 1896 VSSFGVSGTNAHVIVEEAPA----------EQEAARTERGPLPFVLSGRSEAVV 1939
Qy 2556 RAQAQALHAHLTDHPGLDLADVGYTLAHARAVFDHRATLIAADRDTFLQALQALAAGEPH 2615
Db 1940 AAQARALAEHLRDTPELGLTDAAWTLATGRARFDVRAAVLGDDRAGVCAELDALAEGRP- 1998
Qy 2616 PAVIHSSAPGGTGTGEAAGKTAFICSGQGTQRPGMAHGLYHTHPVFAAALNDICTHLDPH 2675
Db 1999 ------SADAVAPVTSAPRKPVLVFPGQGAQWVGMARDLLESSEVFAESMSRCAEALSPH 2052
Qy 2676 LDHPLLPLLTQNDNDNEDAAALLQQTRYAQPALFAFQVALHRLLTDGYHITPHYYAGHSL 2735
Db 2053 TDWKLLDVVRGDGGPDPH-----ERVDVLQPVLFSIMVSLAELWR-AHGVTPAAVVGHSQ 2106
Qy 2736 GEITAAHLAGILTLTDATTLITQRATLMQTM-PPGTMTTLHTTPHHITHHLTAHENDLAI 2794
Db 2107 GEIAAAHVAGALSLEAAAKVVALRSQVLRELDDQGGMVSVGASRDELETVLARWDGRVAV 2166
Qy 2795 AAINTPTSLVISGTPHTVQHITTLCQQQGIKTKTLPTNHAFHSPHTNPILNQLHQHTQTL 2854
Db 2167 AAVNGPGTSVVAGPTAELDEFFAEAEAREMKPRRIAVRYASHSPEVARIEDRLAAELGTI 2226
Qy 2855 TYHPPHTPL---ITANTPPDQLLTPHYWTQQARNTVDYATTTQTLHQHGVTTYIELGPDN 2911
Db 2227 TAVRGSVPLHSTVTGEVIDTSAMDASYWYRNLRRPVLFEQAVRGLVEQGFDTFVEVSPHP 2286
Qy 2912 TLTTLTHHNLPNPPTTTLTLTHPHHHPQTHLLTNLAKTTTTWHPHHYTHHDNQPHTH--- 2968
Db 2287 VLLMAVEE------------TAEHAGAEVTCVPTLRRREQS--GPHEFLRNLLRAHVHGVG 2332
Qy 2969 ----------THLDLPTYPFQHHHYWLESTQPGAGNVSAAGLDPTEHPLLGATLELATDG 3018
Db 2333 ADLRPAVAGGRPAELPTYPFEHQRFWPRPHRP--ADVSALGVRGAEHPLLLAAVDVPGHG 2390
Qy 3019 GALLAGRLSLRSHPWLADHAVGGTVLLSGATFLELALHAGTYVGCDRVDELTLHAPLVVP 3078
Db 2391 GAVFTGRLSTDEQPWLAEHVVGGRTLVPGSVLVDLALAAGEDVGLPVLEELVLQRPLVL- 2449
Qy 3079 VDGGVSVQVGVAAADGEGRRLVSVYARGGSACGGGGASGGVWTCHASGVLVEAAAGGVVV 3138
Db 2450 AGAGALLRMSVGAPDESGRRTIDVHAAEDVA----DLADAQWSQHATGTLAQGVAAG--- 2502
Qy 3139 DGLAGVWPPRGAVAVDVDGVRDRLAGAGCVLGPVFSGLRAVWRDGGDLLAEVCLPEEAWG 3198
Db 2503 PRDTEQWPPEDAVRIPLDDHYDGLAEQGYEYGPSFQALRAAWRKDDSVYAEVSIA----A 2558
Qy 3199 DAAGFGLHPALLDGVVQPLSVLLPGGTGFGEGAGFGEGVRVPAVWGGVSLHRAGVTGVRV 3258
Db 2559 DEEGYAFHPVLLDAVAQTLSLGALGEPG---------GGKLPFAWNTVTLHASGATSVRV 2609
Qy 3259 RVSAVGRGGGREAVSVVVGDEAGVPVASVDRLELRPVDMGQLRAVSVSAGRRGSLYAVQW 3318
Db 2610 ----VATPAGADAMALRVTDPAGHLVATVDSLVVR--STGEKWEQPEPRGGEGELHALDW 2663
Qy 3319 ---AEVGPVPVCGQAWAWHEDVGESGGGPVPGVVVLRC------PDAGAGGGGGGGGGGG 3369
Db 2664 GRLAEPGSTGRVVAADASDLDAVLRSGEPEPDAVLVRYEPEGDDPRAAARHG-------- 2715
Qy 3370 VGEVVGGVLGVVQGWLGLERFAGSRLVVVTRGAVVAGPEDGPVDVVGASVWGLVRSAQAE 3429
Db 2716 ----VLWAAALVRRWLEQEELPGATLVIATSGAVTVSDDDSVPEPGAAAMWGVIRCAQAE 2771
Qy 3430 HPDRFVLLDLD---------- 3440
Db 2772 SPDRFVLLDTDAEPGMLPAVPDNPQLALRGDDVFVPRLSPLAPSALTLPAGTQRLVPGDG 2831
Qy 3441 ----------TDT 3443
Db 2832 AIDSVAFEPAPDVEQPLRAGEVRVDVRATGVNFRDVLLALGMYPQKADMGTEAAGVVTAV 2891
Qy 3444 GTDLDT----------GAGA----------GW---------- 3455
Db 2892 GPDVDAFAPGDRVLGLFQGAFAPIAVTDHRLLARVPDGWSDADAAAVPIAYTTAHYALHD 2951
Qy 3456 ----------GVDGGRVAAVVACGEPQLAVRG----------ERLLAAR 3484
Db 2952 LAGLRAGQSVLIHAAAGGVGMAAVALARRAGAEVLATAGPAKHGTLRALGLDDEHIASSR 3011
Qy 3485 L----------KRLESSGDVPAQ----------RSGDTRAR- 3505
Db 3012 ETGFARKFRERTGGRGVDVVLNSLTGELLDESADLLAEDGVFVEMGKTDLRDAGDFRGRY 3071
Qy 3506 ----------RSDVPAQRSGGVPARRSVDVSGREV----- 3530
Db 3072 APFDLGEAGDDRLGEILREVVGLLGAGELDRLPVSAWELGSAPAALQHMSRGRHVGKLVL 3131
Qy 3531 ---LPWLSGGSVLVTGGTGVLGAAVARHLAGVCGVRDLLLVSRRGPDAPGAEGLRAELAA 3587
Db 3132 TQPAPVDPDGTVLITGGTGTLGRLLARHLVTEHGVRHLLLVSRRGADAPGSDELRAEIED 3191
Qy 3588 LGAEVRIVACDVGERREVVRLLEGVPAGCPLTGVVHAAGVLDDATIASLTPERLGTVFAA 3647
Db 3192 LGASAEIAACDTADRDALSALLDGLPR--PLTGVVHAAGVLADGLVTSIDEPAVEQVLRA 3249
Qy 3648 KVDAALLLDELTRGMELSAFVLFSSAAGILGSAGQGNYAAANAALDALAYRRRAAGLPGV 3707
Db 3250 KVDAAWNLHELTANTGLSFFVLFSSAASVLAGPGQGVYAAANESLNALAALRRTRGLPAK 3309
Qy 3708 SLAWGLWEEASGMTGHLAGTDHRRIIRSGLHPMSTPDALALFDAALALDRPVLLP----- 3762
Db 3310 ALGWGLWAQASEMTSGLGD----RIARTGVAALPTERALALFDSALRRGGEVVFPLSINR 3365
Qy 3763 ADLRPAPPLPPLLQDLLPATRRRTTRTTTTGGAD-NGAQLHARLAGQTHEQQHTTLLALV 3821

```
Db       3366 SALRRAEFVPEVLRGMVRAKLR------AAGQAEAAGPNVVDRLAGRSESDQVAGLAELV 3419

Qy       3822 RSHIATVLGHTTPDTIPPDRAFRDLGFDSLTAVELRNRLSRTTGLRLPTTLAFDHPNPTT 3881

Db       3420 RSHAAAVSGYGSADQLPERKAFKDLGFDSLAAVELRNRLGTATGVRLPSTLVFDHPTPLA 3479

Qy       3882 LTHHLHTQLQPQPDNAVAPVL---AELDKLESALSALDKTDSASERVTLRLKSLMLRWNA 3938

Db       3480 VAEHLRDRL-----FAASPAVDIGDRLDELEKALEALSAEDGHDD-VGQRLESLLRRWNS 3533

Qy       3939 --PQHPTAESADDDEKFTSATEAEIFKFIDNDLG 3970

Db       3534 RRADAPSTSAISED-----ASDDELFSMLDQRFG 3562


RESULT 2
ERY1_SACER
ID   ERY1_SACER     STANDARD;      PRT;  3491 AA.
AC   Q03131;
DT   01-OCT-1993 (Rel. 27, Created)
DT   01-OCT-1993 (Rel. 27, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Erythronolide synthase, modules 1 and 2 (EC 2.3.1.94) (ORF 1) (6-
DE   deoxyerythronolide B synthase I) (DEBS 1).
GN   ERYA.
OS   Saccharopolyspora erythraea (Streptomyces erythraeus).
OC   Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC   Actinomycetales; Pseudonocardineae; Pseudonocardiaceae;
OC   Saccharopolyspora.
OX   NCBI_TaxID=1836;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=91220065; PubMed=2024119;
RA   Donadio S., Staver M.J., McAlpine J.B., Swanson S.J., Katz L.;
RT   "Modular organization of genes required for complex polyketide
RT   biosynthesis.";
RL   Science 252:675-679(1991).
RN   [2]
RP   SEQUENCE OF 3474-3491 FROM N.A.
RX   MEDLINE=93231529; PubMed=8386127;
RA   Donadio S., Staver M.J.;
RT   "IS1136, an insertion element in the erythromycin gene cluster of
RT   Saccharopolyspora erythraea.";
RL   Gene 126:147-151(1993).
CC   -!- CATALYTIC ACTIVITY: 6 malonyl-CoA + propionyl-CoA = 7 CoA + 6-
CC       deoxyerythronolide B.
CC   -!- COFACTOR: NADP; CONTAINS 3 COVALENTLY BOUND PHOSPHOPANTETHEINES.
CC   -!- PATHWAY: COMPLEX POLYKETIDE FORMATION IN ERYTHROMYCIN
CC       BIOSYNTHESIS.
CC   -!- MISCELLANEOUS: IN EACH ORF OF ERYA TWO MODULES ARE PRESENT EACH
CC       ENCODING FOR A FUNCTIONAL SYNTHASE SUBUNIT. THUS ERYA SHOWING 3
CC       ORFS CODES FOR 6 SYNTHASE SUBUNITS. IT IS SUPPOSED THAT EACH
CC       SYNTHASE PARTICIPATES IN ONE OF THE SIX FAS-LIKE ELONGATION STEPS
CC       REQUIRED FOR FORMATION OF THE POLYKETIDE. MODULE 1, 2, 3, 4, 5,
CC       AND 6 PARTICIPATING IN BIOSYNTHESIS STEPS 1, 2, 3, 4, 5, AND 6,
CC       RESPECTIVELY.
CC   -!- MISCELLANEOUS: BIOSYNTHESIS OF POLYKETIDES; ACYLTRANSFERASE (AT),
CC       BETA-KETOACYL CARRIER PROTEIN SYNTHASE (KS), AND ACYL CARRIER
CC       PROTEIN (ACP) FOR CHAIN ELONGATION. BETA-KETOREDUCTASE (KR),
CC       DEHYDRATASE (DH), AND ENOYL REDUCTASE (ER) FOR PROCESSING OF THE
CC       BETA CARBON, AND THIOESTERASE (TE) FOR RELEASE AND LACTONIZATION
CC       OF THE FULL-LENGTH CHAIN.
CC   -!- SIMILARITY: TO FATTY ACID SYNTHASE (FAS).
CC   -!- SIMILARITY: CONTAINS 3 ACYL CARRIER DOMAINS.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; M63676; AAA26493.2; -.
DR   EMBL; L07626; AAA26504.1; -.
DR   HSSP; P25715; 1MLA.
DR   InterPro; IPR002198; ADH_short.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR003880; Ppantne_attach.
DR   Pfam; PF00106; adh_short; 1.
DR   Pfam; PF00109; ketoacyl-synt; 2.
DR   Pfam; PF00550; pp-binding; 3.
DR   Pfam; PF00698; Acyl_transf; 3.
DR   Pfam; PF02801; ketoacyl-synt_C; 2.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; 3.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 2.
DR   PROSITE; PS50075; ACP_DOMAIN; 3.
KW   Transferase; Acyltransferase; Antibiotic biosynthesis; NADP; Repeat;
KW   Phosphopantetheine; Multifunctional enzyme.
FT   DOMAIN        1   1972       MODULE 1.
FT   DOMAIN     1979   3491       MODULE 2.
FT   DOMAIN        1    375       ACYLTRANSFERASE (AT) 1.
FT   DOMAIN      414    484       ACYL CARRIER (ACP) 1.
FT   DOMAIN      503    961       BETA-KETOACYL SYNTHASE 1.
FT   DOMAIN     1030   1356       ACYLTRANSFERASE (AT) 2.
FT   DOMAIN     1611   1794       BETA-KETOACYL REDUCTASE 1.
FT   DOMAIN     1888   1958       ACYL CARRIER (ACP) 2.
FT   DOMAIN     1979   2441       BETA-KETOACYL SYNTHASE 2.
FT   DOMAIN     2507   2854       ACYLTRANSFERASE (AT) 3.
FT   DOMAIN     3055   3237       BETA-KETOACYL REDUCTASE 2.
FT   DOMAIN     3334   3404       ACYL CARRIER (ACP) 3.
FT   ACT_SITE    145    145       ACYL-ENZYME INTERMEDIATE.
FT   BINDING     447    447       PHOSPHOPANTETHEINE (BY SIMILARITY).
FT   ACT_SITE    677    677       BETA-KETOACYL SYNTHASE.
FT   ACT_SITE   1128   1128       ACYL-ENZYME INTERMEDIATE.
FT   NP_BIND    1614   1660       NADP.
FT   BINDING    1921   1921       PHOSPHOPANTETHEINE (BY SIMILARITY).
FT   ACT_SITE   2148   2148       BETA-KETOACYL SYNTHASE.
FT   ACT_SITE   2598   2598       ACYL-ENZYME INTERMEDIATE.
FT   NP_BIND    3058   3104       NADP.
FT   BINDING    3367   3367       PHOSPHOPANTETHEINE (BY SIMILARITY).
SQ   SEQUENCE   3491 AA;  365022 MW;  682BFC32C90FA8C4 CRC64;

  Query Match           30.5%;  Score 6360;  DB 1;  Length 3491;
  Best Local Similarity  39.3%;  Pred. No. 1.1e-281;
  Matches 1588;  Conservative  435;  Mismatches 1294;  Indels  724;  Gaps   89;

Qy         11 RPAAGEVLGVADEADGGVVFVFPGQGPQWPGMGRELLDASDVFRESVRACEAAFAPYVDW 70

Db         40 RPHASVVRGVA-RPSAPVVFVFPGQGAQWAGMAGELLGESRVFAAAMDACARAFEPVTDW 98

Qy         71 SVEQVLRDSPDAPGLDRVDVVQPTLFAVMISLAALWRSQGVEPCAVLGHSLGEIAAAHVS 130

Db         99 TLAQVL-DSPEQS--RRVEVVQPALFAVQTSLAALWRSFGVTPDAVVGHSIGELAAAHVC 155

Qy        131 GGLSLADAARVVTLWSQAQTTLAGTGALVSVAATPDELLPRIAPWTEDNPARLAVAAVNG 190

Db        156 GAAGAADAARAAALWSREMIPLVGNGDMAAVALSADEIEPRIARWDDD----VVLAGVNG 211

Qy        191 PRSTVVSGAREAVADLVADLTAAQVRTRMIPVDVPAHSPLMYAIEERVVSGLLPITPRPS 250

Db        212 PRSVLLTGSPEPVARRVQELSAEGVRAQVINVSMAAHSAQVDDIAEGMRSALAWFAPGGS 271

Qy        251 RIPFHSSVTGGRLDTRELDAAYWYRNMSSTVRFEPAARLLLQQGPKTFVEMSPHPVLTMG 310

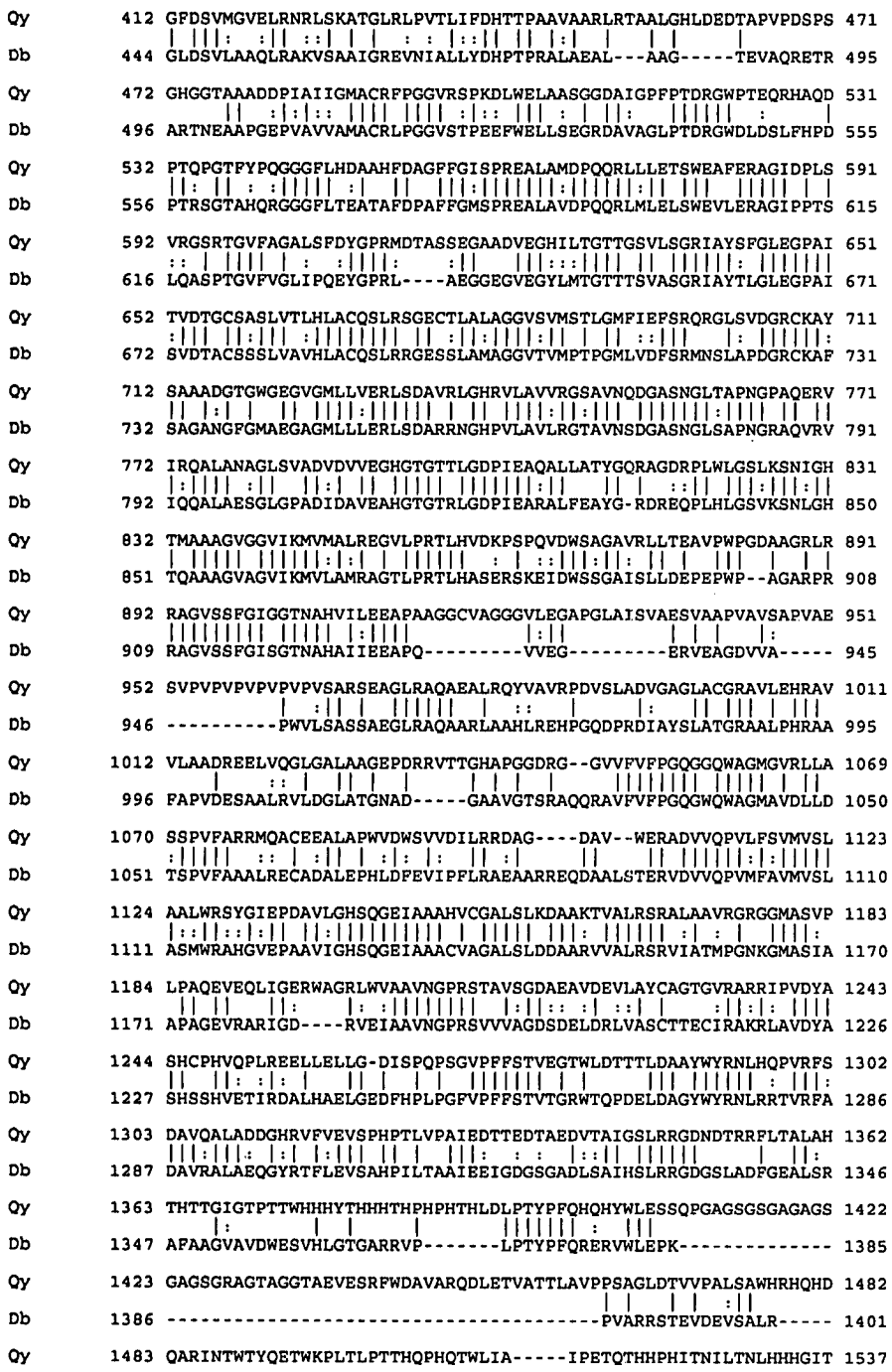
Db        272 EVPFYASLTGGAVDTRELVADYWRRSFRLPVRFDEAIRSALEVGPGTFVEASPHPVLAAA 331

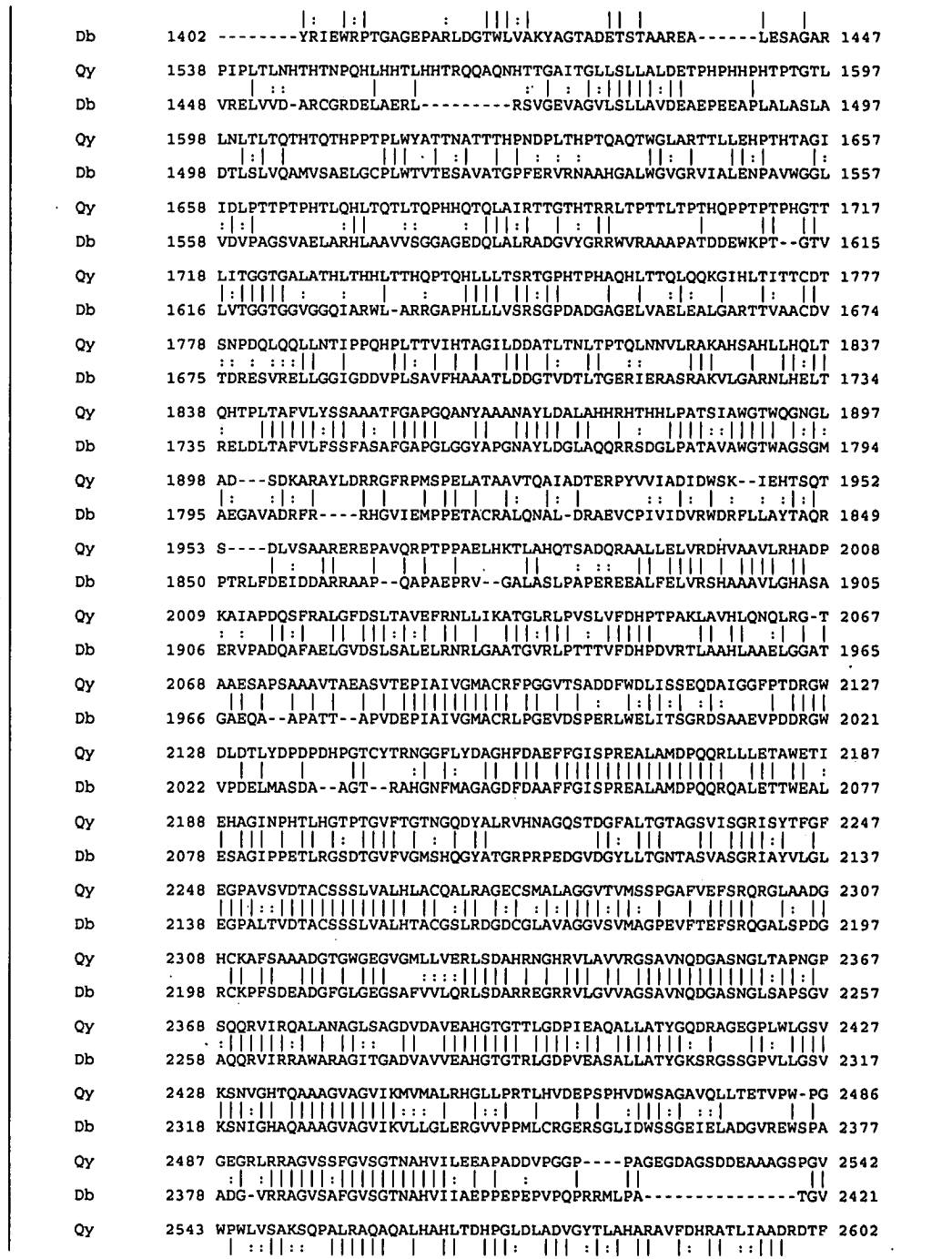
Qy        311 LQELAPDLGDTTGTADTVIMGTLRRGQGTLDHFLTSLAQLRGHGETSATTVLSARLTALS 370

Db        332 LQQTL----DAEGSSAAVV-PTLQRGQGGMRRFLLAAAQAFTGGVAVDWT---AAYDDVG 383

Qy        371 P--------------TQQQSLLLDLVRAHT-----MAVLNDDGNERTASDAGPSASFAHL 411

Db        384 PNPALCRSSRRPRRKTSRPSPASTGTRHRTCCERLLAVVNGETAALAGREADAEATFREL 443
```





```
Db       2422 VPVVLSARTGAALRAQAGRLADHLAAHPGIAPADVSWTMARARQHFEERAAVLAADTAEA 2481

Qy       2603 LQALQALAAGEPHPAVIHSSAPGGTGTGEAAGKTAFICSGQGTQRPGMAHGLYHTHPVFA 2662

Db       2482 VHRLRAVADGAVVPGVVTGSA--------SDGGSVFVFPGQGAQWEGMARELLPV-PVFA 2532

Qy       2663 -------AALNDICTH-----LDPHLDHPLLPLLTQNDNDNEDAAALLQQTRYAQPALFA 2710

Db       2533 ESIAECDAVLSEVAGFSVSEVLEPRPDAP----------------SLERVDVVQPVLFA 2575

Qy       2711 FQVALHRLLTDGYHITPHYYAGHSLGEITAAHLAGILTLTDATTLITQRATLMQTMP-PG 2769

Db       2576 VMVSLARLWR-ACGAVPSAVIGHSQGEIAAAVVAGALSLEDGMRVVARRSRAVRAVAGRG 2634

Qy       2770 TMTTLHTTPHHITHHLT--AHENDLAIAAINTPTSLVISGTPHTVQHITTLCQQQGIKTK 2827

Db       2635 SMLSVRGGRSDVEKLLADDSWTGRLEVAAVNGPDAVVVAGDAQAAREFLEYCEGVGIRAR 2694

Qy       2828 TLPTNHAFHSPHTNPILNQLHQHTQTLTYHPPHTPL---ITANTPPDQLLTPHYWTQQAR 2884

Db       2695 AIPVDYASHTAHVEPVRDELVQALAGITPRRAEVPFFSTLTGDFLDGTELDAGYWYRNLR 2754

Qy       2885 NTVDYATTTQTLHQHGVTTYIELGPDNTLTTLTHHNLPNPPTTTLTL--THPHHHPQTHL 2942

Db       2755 HPVEFHSAVQALTDQGYATFIEVSPHPVLASSVQETLDDAESDAAVLGTLERDAGDADRF 2814

Qy       2943 LTNLAKTTTTWHPHHYTHHDNQPHTHTH---------------LDLPTYPFQHHHYWL-- 2985

Db       2815 LTALA------------------DAHTRGVAVDWEAVLGRAGLVDLPGYPFQGKRFWLLP 2856

Qy       2986 ESTQPGAGNVSAAG----LDPTEHPLLGATLELATDGGALLAGRLSLRSHPWLADHAVGG 3041

Db       2857 DRTTP---RDELDGWFYRVDWTEVP---------RSEPAALRGR-------WL------- 2890

Qy       3042 TVLLSGATFLELALHAGTYVGCDRVDELTLHAPLVVPVDGGVSVQVGVAAADGEGRRLVS 3101

Db       2891 --------------------------------VVVP----------------EG----- 2896

Qy       3102 VYARGGSACGGGGASGGVWTCHASGVLVEAAAGGVVVDGLAGVWPPRGAVAVDVDGVRDR 3161

Db       2897 -HEEDG------------WTVEVRSALAEAGAEPEVTRGVGG------------------ 2925

Qy       3162 LAGAGCVLGPVFSGLRAVWRDGGDLLAEVCLPEEAWGDAAGFGLHPALLDGVVQPLSVLL 3221

Db       2926 ------------------------------------------------------------ 2925

Qy       3222 PGGTGFGEGAGFGEGVRVPAVWGGVSLHRAGVTGVRVRVSAVGRGGGREAVSVVVGDEAG 3281

Db       2926 ---------------------------------------------------LVGDCAG 2932

Qy       3282 VPVASVDRLELRPVDMGQLRAVSVSAGRRGSLYAVQWAEVGPVPVCGQAWAWHEDVGESG 3341

Db       2933 VV--------------------------SLLALE----------------------G 2941

Qy       3342 GGPVPGVVVLRCPDAGAGGGGGGGGGGVGEVVGGVLGVVQGWLGLERFAGSRLVVVTRG 3401

Db       2942 DGAVQTLVLVRELDAE--------------------------GID----APLWTVTFG 2969

Qy       3402 AVVAGPEDGPVDVVGASVWGLVRSAQAEHPDRFV-LLDLDTDTGTDLDTGAGAGWGVDGG 3460

Db       2970 AVDAGSPVARPD--QAKLWGLGQVASLERGPRWTGLVDLPHMPDPEL----------RG 3016

Qy       3461 RVAAVVACGEPQLAVRGERLLAARLKRLESSGDVPAQRSGDTRARRSDVPAQRSGGVPAR 3520

Db       3017 RLTAVLAGSEDQVAVRADAVRARRLS--------PAH----------VTATSEYAVP-- 3055

Qy       3521 RSVDVSGREVLPWLSGGSVLVTGGTGVLGAAVARHLAGVCGVRDLLLVSRRGPDAPGAEG 3580

Db       3056 --------------GGTILVTGGTAGLGAEVARWLAG-RGAEHLALVSRRGPDTEGVGD 3099

Qy       3581 LRAELAALGAEVRIVACDVGERREVVRLLEG-VPAGCPLTGVVHAAGVLDDATIASLTPE 3639

Db       3100 LTAELTRLGARVSVHACDVSSREPVRELVHGLIEQGDVVRGVVHAAGLPQQVAINDMDEA 3159

Qy       3640 RLGTVFAAKVDAALLLDELTRGMELSAFVLFSSAAGILGSAGQGNYAAANAALDALAYRR 3699

Db       3160 AFDEVVAAKAGGAVHLDELCSDAEL--FLLFSSGAGVWGSARQGAYAAGNAFLDAFARHR 3217

Qy       3700 RAAGLPGVSLAWGLWEEASGMTGHLAGTDHRRIIRSGLHPMSTPDALALFDAALALDRPV 3759

Db       3218 RGRGLPATSVAWGLW-AAGGMTGDEEAVSFLR--ERGVRAMPVPRALAALDRVLASGETA 3274

Qy       3760 LLPADL------------RPAPPLPPLLQDLLPATRRRTTRTTTTGGADNGAQLHARLAG 3807

Db       3275 VVVTDVDWPAFAESYTAARPRPLLDRIV----------TTAPSERAGEPETESLRDRLAG 3324

Qy       3808 QTHEQQHTTLLALVRSHIATVLGHTTPDTIPPDRAFRDLGFDSLTAVELRNRLSRTTGLR 3867

Db       3325 LPRAERTAELVRLVRTSTATVLGHDDPKAVRATTPFKELGFDSLAAVRLRNLLNAATGLR 3384

Qy       3868 LPTTLAFDHPNPTTLTHHLHTQLQPQPDNAVAPVLAELDKLESALSALDKTDSASERVTL 3927

Db       3385 LPSTLVFDHPNASAVAGFLDAELGTEVRGEAPSALAGLDALEGALPEVPATE--REELVQ 3442

Qy       3928 RLKSLMLRWNAPQHPTAESAD 3948

Db       3443 RLERML----AALRPVAQAAD 3459
```

```
RESULT 3
OL56_STRAT
ID   OL56_STRAT     STANDARD;      PRT;  3519 AA.
AC   Q07017;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Oleandomycin polyketide synthase, modules 5 and 6.
GN   ORFB.
OS   Streptomyces antibioticus.
OC   Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC   Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=1890;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=94150470; PubMed=8107683;
RA   Swan D.G., Rodriguez A.M., Vilches C., Mendez C., Salas J.A.;
RT   "Characterisation of a Streptomyces antibioticus gene encoding a type
RT   I polyketide synthase which has an unusual coding sequence.";
RL   Mol. Gen. Genet. 242:358-362(1994).
CC   -!- FUNCTION: MAY BE INVOLVED IN THE BIOSYNTHESIS OF THE OLEANDOMYCIN
CC       LACTONE RING.
CC   -!- COFACTOR: CONTAINS 2 COVALENTLY BOUND PHOSPHOPANTETHEINES.
CC   -!- SIMILARITY: CONTAINS 2 ACYL CARRIER DOMAINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L09654; AAA19695.1; -.
DR   HSSP; P25715; 1MLA.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR003880; Ppantne_attach.
DR   InterPro; IPR001031; Thioesterase.
DR   Pfam; PF00109; ketoacyl-synt; 2.
DR   Pfam; PF00550; pp-binding; 2.
DR   Pfam; PF00698; Acyl_transf; 2.
DR   Pfam; PF00975; Thioesterase; 1.
DR   Pfam; PF02801; ketoacyl-synt_C; 2.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; 2.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 2.
DR   PROSITE; PS50075; ACP_DOMAIN; 2.
```

```
KW   Transferase; Acyltransferase; Antibiotic biosynthesis; NADP;
KW   Phosphopantetheine; Multifunctional enzyme; Repeat.
FT   DOMAIN        1      ?       MODULE 5.
FT   DOMAIN        ?   3519       MODULE 6.
FT   DOMAIN       32    501       BETA-KETOACYL SYNTHASE 1.
FT   DOMAIN      569    890       ACYLTRANSFERASE (AT) 1.
FT   DOMAIN     1200   1382       BETA-KETOACYL REDUCTASE 1.
FT   DOMAIN     1487   1561       ACYL CARRIER (ACP) 1.
FT   DOMAIN     1686   2156       BETA-KETOACYL SYNTHASE 2.
FT   DOMAIN     2220   2541       ACYLTRANSFERASE (AT) 2.
FT   DOMAIN     2856   3038       BETA-KETOACYL REDUCTASE 2.
FT   DOMAIN     3141   3215       ACYL CARRIER (ACP) 2.
FT   ACT_SITE      ?      ?       ACYL-ENZYME INTERMEDIATE.
FT   BINDING       ?      ?       PHOSPHOPANTETHEINE (BY SIMILARITY).
FT   DOMAIN     3270   3519       THIOESTERASE.
FT   ACT_SITE    210    210       BETA-KETOACYL SYNTHASE.
FT   ACT_SITE    660    660       ACYL-ENZYME INTERMEDIATE.
FT   NP_BIND    1203   1249       NADP.
FT   BINDING    1524   1524       PHOSPHOPANTETHEINE (BY SIMILARITY).
FT   ACT_SITE   1859   1859       BETA-KETOACYL SYNTHASE.
FT   ACT_SITE   2311   2311       ACYL-ENZYME INTERMEDIATE.
FT   NP_BIND    2859   2905       NADP.
FT   BINDING    3178   3178       PHOSPHOPANTETHEINE (BY SIMILARITY).
SQ   SEQUENCE   3519 AA;  368561 MW;  41AE78AAAEB61F86 CRC64;

  Query Match             29.3%;  Score 6108.5;  DB 1;  Length 3519;
  Best Local Similarity   41.2%;  Pred. No. 3.2e-270;
  Matches 1498;  Conservative  361;  Mismatches 1093;  Indels  687;  Gaps   81;

Qy        479 ADDPIAIIGMACRFPGG----VRSPKDLWELAASGGDAIGPFPTDRGWPTEQRHAQDPTQ 534
              | :||||:||:|||||     | :|:  |:|  |||| |   | ||||    :  :  || :
Db         32 AREPIAIVGMSCRFPGGGDATVNTPEQFWDLLNSGGDGIAGLPEDRGWDLGRLYDPDPDR 91

Qy        535 PGTPYPQGGGFLHDAAHFDAGFFGISPREALAMDPQQRLLLETSWEAFERAGIDPLSVRG 594
              || | : ||||:|:  ||| ||||||||||||||||||||||||||| |||   ::||
Db         92 AGTSYVREGGFLYDSGEFDAAFFGISPREALAMDPQQRLLLETSWEAFESAGIKRAALRG 151

Qy        595 SRTGVFAGALSFDYGPRMDTASS-----EGAADVEGHILTGTTGSVLSGRIAYSFGLEGP 649
              | |||: || |  |       | |     ||   :|| :  |||    |||:||:||||||
Db        152 SDTGVYIGAWSTGY------AGSPYRLVEG---LEGQLAIGTTLGAASGRVAYTFGLEGP 202

Qy        650 AITVDTGCSASLVTLHLACQSLRSGECTLALAGGVSVMSTLGMFIEFSRQRGLSVDGRCK 709
              |:|||| ||:||| |||| | || |||:||| |||:|||:      ||||||||||||||
Db        203 AVTVDTACSSSLVALHLAVQGLRRGECSLALVGGVTVMSSPVTLTTFSRQRGLSVDGRCK 262

Qy        710 AYSAAADGTGWGEGVGMLLVERLSDAVRLGHRVLAVVRGSAVNQDGASNGLTAPNGPAQE 769
              |: |:||| |  ||||:|||||||| ||||||||||||||||||||||||||||||:|:
Db        263 AFPASADGFGAAEGVGVLLVERLSDARRLGHRVLAVVRGSAVNQDGASNGLTAPNGPSQQ 322

Qy        770 RVIRQALANAGLSVADVDVVEGHGTGTTLGDPIEAQALLATYGQ-RAGDRPLWLGSLKSN 828
              |||| |||:|||: ||||||| ||||| ||||||||||||||| ||| ||:||||:|||
Db        323 RVIRAALADAGLAPADVDVVEAHGTGTRLGDPIEAQALLATYGQGRAGGRPVWLGSVKSN 382

Qy        829 IGHTMAAAGVGGVIKMVMALREGVLPRTLHVDKPSPQVDWSAGAVRLLTEAVPWPGDAAG 888
              |||| ||||| ||:|||:||  ||:|:|||||:||| |||||||| |||| ||  :|
Db        383 IGHTQAAAGVAGVMKMVLALGRGVVPKTLHVDEPSPHVDWSAGAVELLTEERPWEPEAE- 441

Qy        889 RLRRAGVSSFGIGGTNAHVILEEAPAAGGCVAGGGVLEGAPGLAISVAESVAAPVAVSAP 948
              ||||||:|:||: |||||||:|||||           |  ||  :  |  :
Db        442 RLRRAGISAFGVSGTNAHVIVEEAPAEP---------EPEPGTRVVAAGDLV-------- 484

Qy        949 VAESVPVPVPVPVPVPVSARSEAGLRAQAEALRQYVAVRPDVSLADVGAGLACGRAVLEH 1008
                          ||  || |     |||||  |  :|:     ||  |||   |   |:| ||
Db        485 -----------VPWVVSGRDARALRAQAARLAAHVS---GVSAVDVGWSLVATRSVFEH 529

Qy       1009 RAVVLAADREELVQGLGALAAGEPDRRVTTGHAPGGDRGGVVFVFPGQGGQWAGMGVRLL 1068
              ||| : :: : :    |  |||     | :| ||   |  |||||||| || ||   ||
Db        530 RAVAIGSELDSMAGSLAGFAAGGVVPGVVSGVAPAEGR-RVVFVFPGQGSQWVGMAAGLL 588

Qy       1069 ASSPVFARRMQACEEALAPWVDWSVVDILR-RDAGDAVWERADVVQPVLFSVMVSLAALW 1127
              : |||| :  |   | |   ||:|::|: |||   |  | ||||| |::|||||  |
Db        589 DACPVFAEAVAECAAVLDPVTGWSLVEVLQGRDA--TVLGRVDVVQPALWAVMVSLARTW 646

Qy       1128 RSYGIEPDAVLGHSQGEIAAAHVCGALSLKDAAKTVALRSRALAAVRGRGGMASVPLPAQ 1187
              | ||:|| ||:|||||||||| | | ||| | |: | |||||:| : | ||| || |||
Db        647 RYYGVEPAAVVGHSQGEIAAACVAGGLSLADGARVVVLRSRAIARIAGGGGMVSVSLPAG 706

Qy       1188 EVEQLIGERWAGRLWVAAVNGPRSTAVSGDAEAVDEVLAYCAGTGVRARRIPVDYASHCP 1247
                |  :: | : ||| ||||||| || |||| :|:||:|| |   ||||||:|||||||
Db        707 RVRTML-EEPDGRLSVAAVNGPSSTVVSGDVQALDELLAGCEREGVRARRVPVDYASHSA 765

Qy       1248 HVQPLREELLELLGDISPQPSGVPFFSTVEGTWLDTTTLDAAYWYRNLHQPVRFSDAVQA 1307
                :  ||:|||| | ||:|| | |||||||   || || | | ||: || : ||| :||:
Db        766 QMDQLRDELLEALADITPQDSSVPFFSTVTADWLGTTALGAGYWFTNLRETVRFQEAVEG 825

Qy       1308 LADDGHRVFVEVSPHPTLVPAIEDTTEDTAEDVTAIGSLRRGDNDTRRFLTALAHTHTTG 1367
              |    |  ||| |||| ||| || | :  ::    ||||| :   ||||:||     |
Db        826 LVAQGMGAFVECSPHPVLVPGIEQTLDALDQNAAVFGSLRRDEGGLDRFLTSLAEAFVQG 885

Qy       1368 IGTPTTWHHHYTHHHTHPHPHTHLDLPTYPFQHQHYWLES-----SQPGAGSGSGAGAGS 1422
              :  | |  :         | | :||||||| ||||| :     |||
Db        886 V--PVDWSRAF----EGVTPRT-VDLPTYPFQRQHYWLMAEEAPVSQP----------- 926

Qy       1423 GAGSGRAGTAGGTAEVESRFWDAVARQDLETVATTLAVPPSAGLDTVVPALSAWHRHQHD 1482
                          |: ||  ||  | |  |  | |   | :: |:||||:|||
Db        927 -------------PHSENSFWSVVADADAEAAAELLGVDVEA-VEAVMPALSSWHRQSQL 972

Qy       1483 QARINTWTYQETWKPLTLPTTHQPHQTWLIAIPETQTHHPHITNILTNLHHHGITPIPLT 1542
              :| :| | |  || ||     :  ||: |       |    :|      : ::
Db        973 RAEVNQWRYDVAWKRLTTGALPEKPGNWLVVTPAGTD-----TTFAESLARTAAAELGVS 1027

Qy       1543 LNHTHTNPQHLHHTLHHTRQQAQNHTTGAITG------LLSLLALDETPHPHHPHTPTGT 1596
              ::    :       | |  | |  :   |:||      |:||||||:         |:
Db       1028 VSFAQVD------TAHPDRSQYAHALRQALTGPENVDHLVSLLALDQAT-DDLAAAPSCL 1080

Qy       1597 LLNLTLTQTHT---QTHPPTPLWYATTNATTTHPND--PLTHPTQAQTWGLARTTLLEHP 1651
                :| | |      :      || |  |    |:|   :  | ||| ||  |   ||||
Db       1081 AASLVLAQALVDLGRVGEGPRLWLVTRGAVVAGPSDAGAVIDPVQAQVWGFGRVLGLEHP 1140

Qy       1652 THTAGIIDLPTTPTPHTLQHLTQTLTQPHHQTQLAIRTTGTHTRRLTPTTLTPTHQPPTP 1711
                |:|||       :      :     : |:|:| :|   |||    :
Db       1141 ELWGGLIDLPVGVDEEVCRRFVGVVASAGFEDQVAVRGSGVWVRRLVRAVV--DGGGGGW 1198

Qy       1712 TPHGTTLITGGTGALATHLTHHLTTHQPTQHLLLTSRTGPHTPHAQHLTTQLQ-QKGIHL 1770
              | || |:||| | |  |    |      |::| || |   | |  | :|:   |  :
Db       1199 RPRGTVLVTGGLGGLGAHTARWL-VGGGADHVVLVSRRGGSAPGAGDLVRELEGLGGARV 1257

Qy       1771 TITTCDTSNPDQLQQLLNTIPPQHPLTTVIHTAGILDDATLTNLTPTQLNNVLRAKAHSA 1830
              ::  || ::   |: ||: :    |:| | | ||:     |  ::  :  :|: ||   |
Db       1258 SVRACDVADRVALRALLSDL--GEPVTAVFHAAGVPQSTPLAEISVQEAADVMAAKVAGA 1315

Qy       1831 HLLHQLTQHTPLTAFVLYSSAAATFGAPGQANYAAANAYLDALAHHRHTHHLPATSIAWG 1890
              | :|     | ||||:|| |  :|: ||| ||||||:|||||  |    |||||:|||
Db       1316 VNLGELVDPCGLEAFVLFSSNAGVWGSGGQAVYAAANAFLDALAVRRRGVGLPATSVAWG 1375

Qy       1891 TWQGNGLADSDKARAYLDRRGFRPMSPELATAAVTQAIADTERPYVVIADIDWSKIE--- 1947
              | | |:|    |   | ||| | | || | | :  |:   |  :| :||:|| :
Db       1376 MWAGEGMASVGGAARELSRRGVRAMDPERAVAVMADAVGRGE-AFVAVADVDWERFVTGF 1434

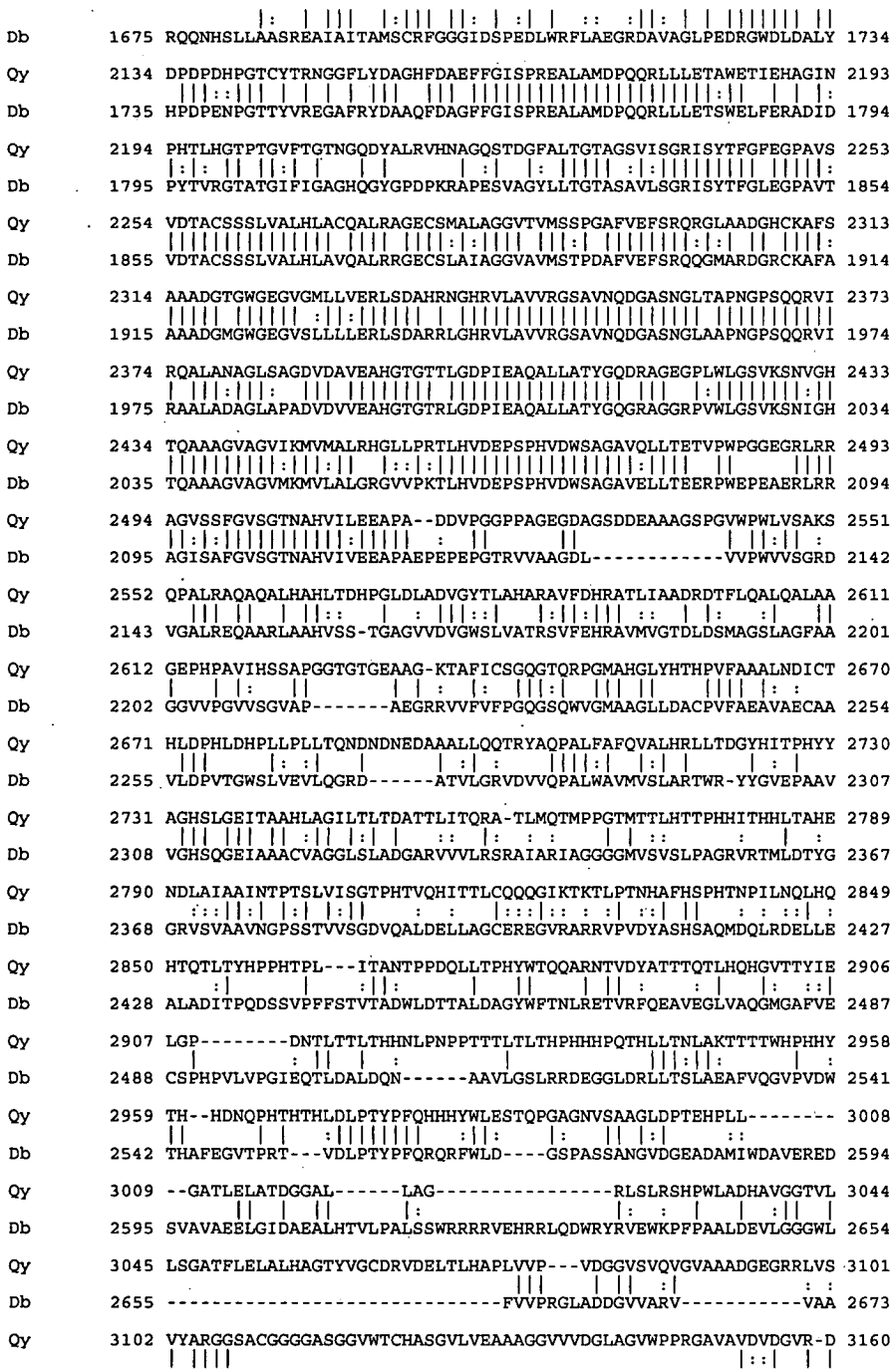
Qy       1948 HTSQTSDLVSAARAREPAVQRPT-------------PPAELHKTLAHQTSADQRAALLEL 1994
              :::   |:|   |   |:                    :   | |:  :   |  |:||
Db       1435 ASARPRPLISDLPEVRAVVEGQVQGRGQGLGLVGEEESSGWLKRLSGLSRVRQEEELVEL 1494

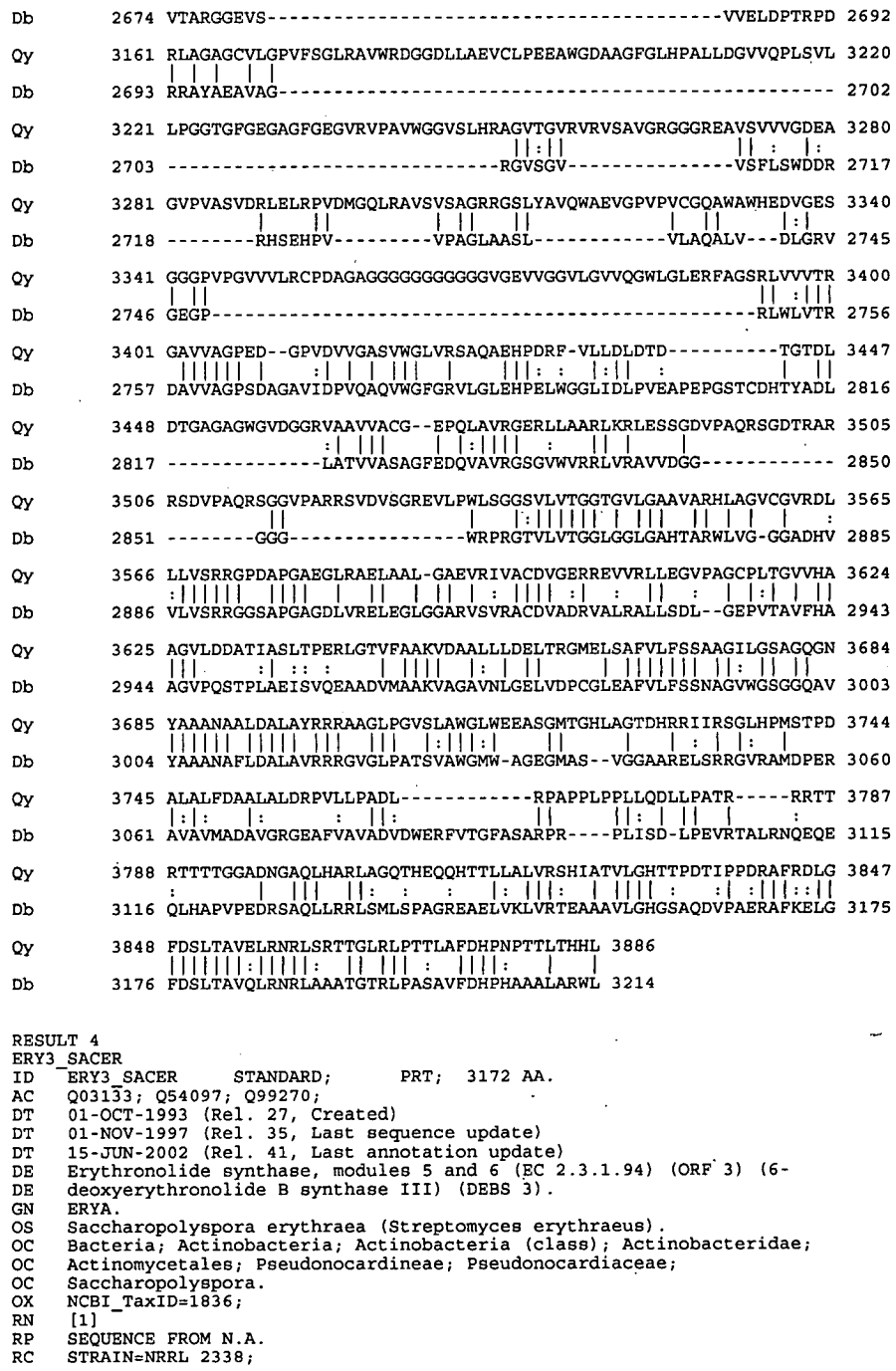
Qy       1995 VRDHVAAVLRHADPKAIAPDQSFRALGFDSLTAVEFRNLLIKATGLRLPVSLVFDHPTPA 2054
              ||   | || |   : :  :::|: ||||||||| || |  |||:||| :: ||||
Db       1495 VRAQAAVVLGHGSAQDVPAERAFKELGFDSLTAVELRNGLAAATGIRLPATMAFDHPNAT 1554

Qy       2055 KLAVHLQNQLRGTA-AESA-PS---------------------------------------- 2074
              :|  ||:||   | :||| ||
Db       1555 AIARFLQSQLLPDAESESAVPSSPEDEVRQALASLSLDQLKGAGLLDPLLALTRLREINS 1614

Qy       2075 ----------------------AAAVTAE---------------------------- 2081
                                     :|  |||
Db       1615 TVQNPEPTTESIDEMDGETCCAWRSAKSTAEPLTTGADMPDPTAKYVEALRASLKENERL 1674

Qy       2082 --------ASVTEPIAIVGMACRFPGGVTSADDFWDLISSEQDAIGGFPTDRGWDLDTLY 2133
```





RESULT 4
ERY3_SACER
ID ERY3_SACER STANDARD; PRT; 3172 AA.
AC Q03133; Q54097; Q99270;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Erythronolide synthase, modules 5 and 6 (EC 2.3.1.94) (ORF 3) (6-
DE deoxyerythronolide B synthase III) (DEBS 3).
GN ERYA.
OS Saccharopolyspora erythraea (Streptomyces erythraeus).
OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC Actinomycetales; Pseudonocardineae; Pseudonocardiaceae;
OC Saccharopolyspora.
OX NCBI_TaxID=1836;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NRRL 2338;

```
RX   MEDLINE=91043075; PubMed=2234082;
RA   Cortes J., Haydock S.F., Roberts G.A., Bevitt D.J., Leadlay P.F.;
RT   "An unusually large multifunctional polypeptide in the erythromycin-
RT   producing polyketide synthase of Saccharopolyspora erythraea.";
RL   Nature 348:176-178(1990).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=91220065; PubMed=2024119;
RA   Donadio S., Staver M.J., McAlpine J.B., Swanson S.J., Katz L.;
RT   "Modular organization of genes required for complex polyketide
RT   biosynthesis.";
RL   Science 252:675-679(1991).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NRRL 2338;
RX   MEDLINE=92155230; PubMed=1740151;
RA   Bevitt D.J., Cortes J., Haydock S.F., Leadlay P.F.;
RT   "6-Deoxyerythronolide-B synthase 2 from Saccharopolyspora erythraea.
RT   Cloning of the structural gene, sequence analysis and inferred domain
RT   structure of the multifunctional enzyme.";
RL   Eur. J. Biochem. 204:39-49(1992).
CC   -!- CATALYTIC ACTIVITY: 6 malonyl-CoA + propionyl-CoA = 7 CoA + 6-
CC       deoxyerythronolide B.
CC   -!- COFACTOR: NADP; CONTAINS 2 COVALENTLY BOUND PHOSPHOPANTETHEINES.
CC   -!- PATHWAY: COMPLEX POLYKETIDE FORMATION IN ERYTHROMYCIN
CC       BIOSYNTHESIS.
CC   -!- MISCELLANEOUS: IN EACH ORF OF ERYA TWO MODULES ARE PRESENT EACH
CC       ENCODING FOR A FUNCTIONAL SYNTHASE SUBUNIT. THUS ERYA SHOWING 3
CC       ORFS CODES FOR 6 SYNTHASE SUBUNITS. IT IS SUPPOSED THAT EACH
CC       SYNTHASE PARTICIPATES IN ONE OF THE SIX FAS-LIKE ELONGATION STEPS
CC       REQUIRED FOR FORMATION OF THE POLYKETIDE. MODULE 1, 2, 3, 4, 5,
CC       AND 6 PARTICIPATING IN BIOSYNTHESIS STEPS 1, 2, 3, 4, 5, AND 6,
CC       RESPECTIVELY.
CC   -!- MISCELLANEOUS: BIOSYNTHESIS OF POLYKETIDES; ACYLTRANSFERASE (AT),
CC       BETA-KETOACYL CARRIER PROTEIN SYNTHASE (KS), AND ACYL CARRIER
CC       PROTEIN (ACP) FOR CHAIN ELONGATION. BETA-KETOREDUCTASE (KR),
CC       DEHYDRATASE (DH), AND ENOYL REDUCTASE (ER) FOR PROCESSING OF THE
CC       BETA CARBON, AND THIOESTERASE (TE) FOR RELEASE AND LACTONIZATION
CC       OF THE FULL-LENGTH CHAIN.
CC   -!- SIMILARITY: TO FATTY ACID SYNTHASE (FAS).
CC   -!- SIMILARITY: CONTAINS 2 ACYL CARRIER DOMAINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X56107; CAA39583.1; -.
DR   EMBL; M63677; AAA26495.1; -.
DR   EMBL; X62569; CAA44449.1; -.
DR   HSSP; P00101; 1CCH.
DR   InterPro; IPR002198; ADH_short.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR003880; Ppantne_attach.
DR   InterPro; IPR001031; Thioesterase.
DR   Pfam; PF00106; adh_short; 1.
DR   Pfam; PF00109; ketoacyl-synt; 2.
DR   Pfam; PF00550; pp-binding; 2.
DR   Pfam; PF00698; Acyl_transf; 2.
DR   Pfam; PF00975; Thioesterase; 1.
DR   Pfam; PF02801; ketoacyl-synt_C; 2.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; 2.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 2.
DR   PROSITE; PS50075; ACP_DOMAIN; 2.
KW   Transferase; Acyltransferase; Antibiotic biosynthesis; NADP; Repeat;
KW   Phosphopantetheine; Multifunctional enzyme.
FT   DOMAIN        1   1484       MODULE 5.
FT   DOMAIN     1485   3172       MODULE 6.
FT   DOMAIN       37    484       BETA-KETOACYL SYNTHASE 1.
FT   DOMAIN      554    878       ACYLTRANSFERASE (AT) 1.
FT   DOMAIN     1116   1298       BETA-KETOACYL REDUCTASE 1.
FT   DOMAIN     1394   1464       ACYL CARRIER (ACP) 1.
FT   DOMAIN     1488   1954       BETA-KETOACYL SYNTHASE 2.
FT   DOMAIN     2021   2335       ACYLTRANSFERASE (AT) 2.
FT   DOMAIN     2555   2735       BETA-KETOACYL REDUCTASE 2.
FT   DOMAIN     2821   2891       ACYL CARRIER (ACP) 2.
FT   DOMAIN     2926   3172       THIOESTERASE.
FT   ACT_SITE    199    199       THIOESTER BOND.
FT   ACT_SITE    643    643       ACYL-ENZYME INTERMEDIATE.
FT   NP_BIND    1118   1164       NADP.
FT   BINDING    1427   1427       PHOSPHOPANTETHEINE (BY SIMILARITY).
FT   ACT_SITE   1661   1661       THIOESTER BOND.
FT   ACT_SITE   2112   2112       ACYL-ENZYME INTERMEDIATE.
FT   NP_BIND    2557   2605       NADP.
FT   BINDING    2854   2854       PHOSPHOPANTETHEINE (BY SIMILARITY).
FT   CONFLICT    231    231       S -> SA (IN REF. 2).
FT   CONFLICT    240    240       MISSING (IN REF. 2).
FT   CONFLICT    289    289       A -> R (IN REF. 2).
FT   CONFLICT    493    493       P -> R (IN REF. 2).
FT   CONFLICT    493    517       PEPRNSLRDTGFTLATRASAMEHRA -> ASRGTRCATPVS
FT                                RWPPAAAPWEQ (IN REF. 1).
FT   CONFLICT    510    510       A -> R (IN REF. 2).
FT   CONFLICT    513    513       M -> W (IN REF. 2).
FT   CONFLICT    525    525       E -> D (IN REF. 2).
FT   CONFLICT    536    536       R -> G (IN REF. 2).
FT   CONFLICT    547    551       GPNSP -> ARTR (IN REF. 2).
FT   CONFLICT    553    553       R -> G (IN REF. 2).
FT   CONFLICT    673    673       R -> A (IN REF. 2).
FT   CONFLICT    716    716       MISSING (IN REF. 2).
FT   CONFLICT    734    736       AHK -> GIT (IN REF. 2).
FT   CONFLICT    896    896       R -> RQR (IN REF. 2).
FT   CONFLICT    896    896       R -> RELPVYPFQRQR (IN REF. 1).
FT   CONFLICT    988    994       GVAAVPH -> VSLLSRD (IN REF. 2).
FT   CONFLICT   1108   1116       RTHPLEPLA -> ARTRWSPR (IN REF. 2).
FT   CONFLICT   1124   1126       MISSING (IN REF. 1).
FT   CONFLICT   1132   1132       L -> V (IN REF. 2).
FT   CONFLICT   1192   1192       A -> R (IN REF. 2).
FT   CONFLICT   1194   1194       MISSING (IN REF. 2).
FT   CONFLICT   1277   1278       AA -> RR (IN REF. 2).
FT   CONFLICT   1385   1390       LCDGRE -> STAER (IN REF. 2).
FT   CONFLICT   1485   1485       MISSING (IN REF. 2).
FT   CONFLICT   1518   1518       G -> R (IN REF. 2).
FT   CONFLICT   1601   1601       V -> L (IN REF. 2).
FT   CONFLICT   1724   1725       LP -> FA (IN REF. 2).
FT   CONFLICT   1732   1732       Q -> L (IN REF. 2).
FT   CONFLICT   1739   1743       GPAEG -> ARRA (IN REF. 2).
FT   CONFLICT   1762   1762       T -> S (IN REF. 2).
FT   CONFLICT   2252   2252       D -> DGAD (IN REF. 2).
FT   CONFLICT   2275   2277       QSP -> AVA (IN REF. 2).
FT   CONFLICT   2408   2408       G -> GR (IN REF. 2).
FT   CONFLICT   2420   2421       LA -> S (IN REF. 2).
FT   CONFLICT   2443   2444       NA -> TH (IN REF. 2).
FT   CONFLICT   2596   2596       A -> G (IN REF. 2).
FT   CONFLICT   2609   2609       P -> A (IN REF. 2).
FT   CONFLICT   2715   2722       RRAEGRAA -> AVRKAVRR (IN REF. 1).
FT   CONFLICT   2754   2754       D -> E (IN REF. 2).
SQ   SEQUENCE   3172 AA;  331474 MW;  DBBD5094E77DDD5F CRC64;

  Query Match             25.9%;  Score 5392;  DB 1;  Length 3172;
  Best Local Similarity   38.5%;  Pred. No. 1.2e-237;
  Matches 1343;  Conservative  357;  Mismatches 1082;  Indels  706;  Gaps   77;

Qy        479 ADDPIAIIGMACRFPGGVRSPKDLWELAASGGDAIGPFPTDRGWPTEQRHAQDPTQPGTF 538
              | :||||:|||||||| | ||:  ||  : |||||   | ||||        ||
Db         37 AGEPIAIVGMACRFPGDVDSPESFWEFVSGGGDAIAEAPADRGW------EPDPD----- 85

Qy        539 YPQGGGFLHDAAHFDAGFFGISPREALAMDPQQRLLLETSWEAFERAGIDPLSVRGSRTG 598
                : || |  |  |||||||||||||||||||||::|| |||| |||| ||:|:||| ||
Db         86 -ARLGGMLAAAGDFDAGFFGISPREALAMDPQQRIMLEISWEALERAGHDPVSLRGSATG 144

Qy        599 VFAGALSFDYGPRMDTASSEGAADVEGHILTGTTGSVLSGRIAYSFGLEGPAITVDTGCS 658
              || |  : ||||| | |  |    | |:: |||  || |||:||  ||||||:|||| ||
```



```
Qy       2787 AHENDLAIAAINTPTSLVISGTPHTVQHITTLCQQQGIKTKTLPTNHAFHSPHTNPILNQ 2846
Db       2166 PWEDRISVAAVNSPSSVVVSGDPEALAELVARCEDEGVRAKTLPVDYASHSRHVEEIRET 2225

Qy       2847 LHQHTQTLTYHPPHTPLITANTPPDQLLTPHYWTQQARNTVDYATTTQTLHQHGVTTYIE 2906
Db       2226 ILADLDGISARRAAIPLYSTLHGERRDMGPRYWYDNLRSQVRFDEAVSAQSPDGHATFVE 2285

Qy       2907 LGPDNTLTTLTHHNLPNPPTTTLTLTHPHHH-PQTHLLTNLAKTTTTWHPHHYTHHDNQP 2965
Db       2286 MSPHPVLTAAVQEIAAD----AVAIGSLHRDTAEEHLIAELAR---------------- 2324

Qy       2966 HTHTH---------------LDLPTYPPQHHHYWLESTQPGAGNVSAAGLDPTEHPLLGA 3010
Db       2325 -AHVHGVAVDWRNVFPAAPPVALPNYPFEPQRYWL------------------------ 2358

Qy       3011 TLELATDGGALLAGRLSLRSHPWLADHAVGGTVLLSGATFLELALHAGTYVGCDRVDELT 3070
Db       2359 ------------------------------------------------------------ 2358

Qy       3071 LHAPLVVPVDGGVSVQVGVAAADGEGRRLVSVYARGGSACGGGGASGGVWTCHASGVLVE 3130
Db       2359 --AP-----------EVSDQLADSRYR------------------------------- 2372

Qy       3131 AAAGGVVVDGLAGVWPPRGAVAVDVDGVRDRLAGAGCVLGPVFSGLRAVWRDGGDLLAEV 3190
Db       2373 -------VD-----WRPLATTPVDLEG------------------------------- 2387

Qy       3191 CLPEEAWGDAAGFGLHPALLDGVVQPLSVLLPGGTGFGEGAGFGEGVRVPAVWGGVSLHR 3250
Db       2388 -----------GFLVHGSAPESL---------------------------------- 2399

Qy       3251 AGVTGVRVRVSAVGRGGGREAVSVVVGDEAGVPVASVDR----LELRPVDMGQLRAVSVS 3306
Db       2400 ---------TSAVEKAGG------------VVPVASADREALAAALREVPGEVAGVLSVH 2438

Qy       3307 AGRRGSLYAVQWAEVGPVPVCGQAWAWHEDVGESGGGPVPGVVVLRCPDAGAGGGGGGGG 3366
Db       2439 TGAANAL-----------------ALHQSLGEAG---------VRAP------------ 2459

Qy       3367 GGGVGEVVGGVLGVVQGWLGLERFAGSRLVVVTRGAVVAGPEDGPVDVVGASVWGLVRSA 3426
Db       2460 --------------------------LWLVTSRAVALG-ESEPVDPEQAMVWGLGRVM 2490

Qy       3427 QAEHPDRF-VLLDLDTDTGTDLDTGAGAGWGVDGGRVAAVVAC-----GEPQLAVR---- 3476
Db       2491 GLETPERWGGLVDLPAEPAP-------------GDGEAFVACLGADGHEDQVAIRDHAR 2536

Qy       3477 -GERLLAARLKRLESSGDVPAQRSGDTRARRSDVPAQRSGGVPARRSVDVSGREVLPWLS 3535
Db       2537 YGRRLVRAPLGTRESS-------------------------------------------WEP 2555

Qy       3536 GGSVLVTGGTGVLGAAVARHLAGVCGVRDLLLVSRRGPDAPGAEGLRAELAALGAEVRIV 3595
Db       2556 AGTALVTGGTGALGGHVARHLAR-CGVEDLVLVSRRGVDAPAAAELEAELVALGPKTTIT 2614

Qy       3596 ACDVGERREVVRLLEGVPA-GCPLTGVVHAAGVLDDATIASLTPERLGTVFAAKVDAALL 3654
Db       2615 ACDVADREQLSKLLEELRGQGRPVRTVVHTAGVPESRPLHEI--GELESVCAAKVTGARL 2672

Qy       3655 LDELTRGMELSAFVLFSSAAGILGSAGQGNYAAANAALDALAYRRRAAGLPGVSLAWGLW 3714
Db       2673 LDELCPDAE--TFVLFSSGAGVWGSANLGAYSAANAYLDALAHRRRAEGRAATSVAWGAW 2730

Qy       3715 EEASGMTGHLAGTDHRRIIRSGLHPMSTPDALALFDAALALDRPVLLPADL--------- 3765
Db       2731 AGEGMATGDLEG-----LTRRGLRPMAPDRAIRALHQALDNGDTCVSIADVDWEAFAVGF 2785

Qy       3766 ---RPAPPLPPLLQDLLPATRRRTTRTTTTGGADNGAQLHARLAGQTHEQQHTTLLALVR 3822
Db       2786 TAARPR----PLLDELV----------TPAVGAVPAVQ-----AAPAREMTSQELLEFTH 2826

Qy       3823 SHIATVLGHTTPDTIPPDRAFRDLGFDSLTAVELRNRLSRTTGLRLPTTLAFDHPNPTTL 3882
Db       2827 SHVAAILGHSSPDAVGQDQPFTELGFDSLTAVGLRNQLQQATGLALPATLVFEHPTVRRL 2886

Qy       3883 THHLHTQL 3890
Db       2887 ADHIGQQL 2894
```

```
RESULT 5
PPSB_MYCTU
ID   PPSB_MYCTU     STANDARD;      PRT;  1538 AA.
AC   Q10978; O53234;
DT   01-OCT-1996 (Rel. 34, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Phenolpthiocerol synthesis polyketide synthase ppsB.
GN   PPSB OR RV2932 OR MT3002 OR MTCY338.21 OR MTV011.01.
OS   Mycobacterium tuberculosis.
OC   Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC   Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1773;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H37Rv;
RX   MEDLINE=98295987; PubMed=9634230;
RA   Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA   Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekaia F.,
RA   Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA   Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA   Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA   Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA   Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA   Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT   "Deciphering the biology of Mycobacterium tuberculosis from the
RT   complete genome sequence.";
RL   Nature 393:537-544(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CDC 1551 / Oshkosh;
RA   Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA   Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA   Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA   Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA   Bishai W.;
RT   "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT   laboratory strains.";
RL   Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: POLYKETIDE SYNTHASE INVOLVED IN PHENOLPTHIOCEROL
CC       SYNTHESIS.
CC   -!- COFACTOR: CONTAINS 1 COVALENTLY BOUND PHOSPHOPANTETHEINE (By
CC       similarity).
CC   -!- SIMILARITY: CONTAINS 1 ACYL CARRIER DOMAIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AL021070; CAA15929.1; -.
DR   EMBL; AE007122; AAK47329.1; -.
DR   TIGR; MT3002; -.
DR   TubercuList; Rv2932; -.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR003880; Ppantne_attach.
DR   Pfam; PF00109; ketoacyl-synt; 1.
DR   Pfam; PF00550; pp-binding; 1.
DR   Pfam; PF00698; Acyl_transf; 1.
DR   Pfam; PF02801; ketoacyl-synt_C; 1.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 1.
```

```
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; FALSE_NEG.
DR   PROSITE; PS50075; ACP_DOMAIN; 1.
KW   Hypothetical protein; Multifunctional enzyme; Oxidoreductase;
KW   Transferase; NADP; Phosphopantetheine; Complete proteome.
FT   DOMAIN     1425   1495       ACYL CARRIER (ACP).
FT   BINDING    1458   1458       PHOSPHOPANTETHEINE (POTENTIAL).
SQ   SEQUENCE   1538 AA;  162527 MW;  B55E2A2042AD00CC CRC64;

  Query Match             9.8%;  Score 2050.5;  DB 1;  Length 1538;
  Best Local Similarity   33.5%;  Pred. No. 6.4e-86;
  Matches  558;  Conservative  233;  Mismatches  656;  Indels  219;  Gaps   44;

Qy        478 AADDPIAIIGMACRFPGGVRSPKDLWELAASGGDAIGPFPTDRGWPTEQRHAQDPTQPGT 537
Db         31 AVAEPVAVVGIGCRFPGDVDGPESFWDFLVAGRNAISTVPADR-WDAEAFYHPDPLTPGR  89

Qy        538 FYPQGGGFLHDAAHFDAGFFGISPREALAMDPQQRLLLETSWEAFERAGIDPLSVRGSRT 597
Db         90 MTTKWGGFVPDVAGFDAEFFGITPREAAAMDPQQRMLLEVAWEALEHAGIPPDSLGGTRT 149

Qy        598 GVFAGALSFDYGPRMDTASSEGAADVEGHILTGTTGSVLSGRIAYSFGLEGPAITVDTGC 657
Db        150 AVMMGVYFNEY-QSMLAASPQ---NVDAYSGTGNAHSITVGRISYLLGLRGPAVAVDTAC 205

Qy        658 SASLVTLHLACQSLRSGECTLALAGGVSVMSTLGMFIEFSRQRGLSVDGRCKAYSAAADG 717
Db        206 SSSLVAVHLACQSLRLRETDLALAGGVSITLRPETQIAISAWGLLSPQGRCAAFDAAADG 265

Qy        718 TGWGEGVGMLLVERLSDAVRLGHRVLAVVRGSAVNQDGASNGLTAPNGPAQERVIRQALA 777
Db        266 FVRGEGAGVVVLKRLTDAVRDGDQVLAVVRGSAVNQDGRSNGVTAPNTAAQCDVIADALR 325

Qy        778 NAGLSVADVDVVEGHGTGTTLGDPIEAQALLATYGQRAGDRPLWLGSLKSNIGHTMAAAG 837
Db        326 SGDVAPDSVNYVEAHGTGTVLGDPIEFEALAATYGH--GGDACALGAVKTNIGHLEAAAG 383

Qy        838 VGGVIKMVMALREGVLPRTLHVDKPSPQVDWSAGAVRLLTEAVPWPGDAAGRLRRAGVSS 897
Db        384 IAGFIKATLAVQRATIPPNLHFSQWNPAIDAASTRFFVPTQNSPWP--TAEGPRRAAVSS 441

Qy        898 FGIGGTNAHVILEE----APAAGGCVAGGGVLEGAPGLAIS--VAESVAAPVAVSAPVAE 951
Db        442 FGLGGTNAHVIIEQGSELAP-----VSEGGEDTGVSTLVVTGKTAQRMAATAQVLADWME 496

Qy        952 SVPVPVPVPVPVPVSARSEAGLRAQAEALRQYVAVRPDVSLADVGAGLACGRAVLEHRA- 1010
Db        497 GPGA--------------------------------EVAVADV------AHTVNHHRAR 517

Qy       1011 -----VVLAADREELVQGLGALAAGEPDRRVTTGHAPG------GDRG-GVVFVFPGQGG 1058
Db        518 QATFGTVVARDRAQAIAGLRALAAGQ--------HAPGVVSHQDGSPGPGTVFVYSGRGS 569

Qy       1059 QWAGMGVRLLASSPVFARRMQACEEALAPWVDWSVVDILRRDAGDAVWERADVVQPVLFS 1118
Db        570 QWAGMGRQLLADEPAFAAAVAELEPVFVEQAGFSLRDVI---ATGKELVGIEQIQLGLIG 626

Qy       1119 VMVSLAALWRSYGIEPDAVLGHSQGEIAAAHVCGALSLKDAAKTVALRSRALAAVRGRGG 1178
Db        627 MQLTLTELWRSYGVQPDLVIGHSMGEVAAAVVAGALTPAEGLRVTATRARLMAPLSGQGG 686

Qy       1179 MASVPLPAQEVEQLIGERWAGRLWVAAVNGPRSTAVSGDAEAVDEVLAYCAGTGVRARRI 1238
Db        687 MALLGLDAAATEALIAD--YPQVTVGIYNSPRQTVIAGPTEQIDELIARVRAQNRFASRV 744

Qy       1239 PVDYASHCPHVQPLREELLELLGDISPQPSGVPFFSTVEGTWLDTTT---LDAAYWYRNL 1295
Db        745 NIEVAPHNPAMDALQPAMRSELADLTPRTPTIGIIST---TYADLHTQPIFDAEHWATNM 801

Qy       1296 HQPVRFSDAVQAL---ADDGHRVFVEVSPHPTLVPAIEDTTED------TAEDVTAIGSL 1346
Db        802 RNPVRFQQAIASAGSGADGAYHTFIEISAHPLLTQAIADTLEDAHRPTKSAAKYLSIGTL 861

Qy       1347 RRGDNDTRRFLTALAHTHTTGIGTPTTWHHHYTHHHTHPHPHTHLDLPTYPFQHQHYWLE 1406
Db        862 QRDADDTVTFRTNL---YTADIAHPP--------HTCHP-PEPHPTIPTTPWQHTHHWIA 909

Qy       1407 SSQPGAGSGSGAGAGSGAGSGRAGTAGGTAEVESRFWDAVARQDLETVATTLAVPPSAGL 1466
Db        910 TTHPSTAAPEDPGSNKVVVNGQSTS-------ESR--------ALEDWCHQLAWPIRPAV 954

Qy       1467 DTVVPALSAW-----HRHQHDQARINTWTYQETWKPLTLPTTHQPHQTWLIAIPETQTHH 1521
Db        955 SADPPSTAAWLVVADNELCHELARAAD-SRVDSLSPPALAAGSDP-AALLDAL------- 1005

Qy       1522 PHITNILTNLHHHGITPIPLTLNHTHTNPQHLHHTLHHTRQQAQNHTTGAITGLLSLLAL 1581
Db        1006 RGVDNVLY------APPVPGEL----LDIESAYQVFHATRRLAAAMVASSATAI------ 1049

Qy       1582 DETPHPHHPHTPTGTLLNLTLTQTHTQTHPPTPLWYATTNATTTHPNDPLTHPTQAQTWG 1641
Db        1050 ----------------------------SPPKLFIMTRNAQPISEGD-RANPGHAVLWG 1079

Qy       1642 LARTTLLEHPTHTAGIIDL-PTTPTPHTLQHLTQTLTQPH---HQTQLAIRTTGTHTRRL 1697
Db        1080 LGRSLALEHPEIWGGIIDLDDSMPAELAVRHV---LTAAHGTDGEDQVVYRSGARHVPRL 1136

Qy       1698 TPTTLTPTHQPPTPTPHGTTLITGGTGALATHLTHHLTTHQPTQHLLLTSRTGPHTPHAQ 1757
Db        1137 QRRTL--PGKPVTLNADASQLVIGATGNIGPHLIRQLARMGAKTIVAMARKPGA----LD 1190

Qy       1758 HLTTQLQQKGIHLTITTCDTSNPDQLQQLLNTIPPQHPLTTVIHTAGILD-DATLTNLTP 1816
Db        1191 ELTQCLAATGTDLIAVAADATDPAAMQTLFDRFGTELPPLEGIYLAAFAGRPALLSEMTD 1250

Qy       1817 TQLNNVLRAKAHSAHLLHQLTQHTPLTAFVLYSSAAATFGAPGQANYAAANAYLDALAHH 1876
Db        1251 DDVTTMFRPKLDALALLHRRSLKSPVRHFVLFSSVSGLLGSRWLAHYTATSAFLDSFAGA 1310

Qy       1877 RHTHHLPATSIAWGTWQGNGLADSDKARAYLD-RRGFRPMSPELATAAVTQAIADTERPY 1935
Db        1311 RRTMGLPATVVDWGLWK--SLADVQKDATQISAESGLQPMADEVAIGALPLVMNPDAAVA 1368

Qy       1936 VVIADIDWSKIEHTSQTSDLVSAARAREPAVQRPTPPAELHKTLAHQTSADQRAALLELV 1995
Db        1369 TVVVAADWPLLAAAYRTRGALRIVDDLLPAPEDVGKGESEFRTSLRSCPAEKRR---DML 1425

Qy       1996 RDHV----AAVLRHADPKAIAPDQSFRALGFDSLTAVEFRNLLIKATGLRLPVSLVFDHP 2051
Db        1426 FDHVGALAATVMGMPPTEPLDPSAGFFQLGMDSLMSVTLQRALSESLGEFLPASVVFDYP 1485

Qy       2052 TPAKLAVHLQNQL--------RGTAAESAPSAAAVTAEASVTEPIA 2089
Db        1486 TVYSLTDYLATVLPELLEIGATAVATQQATDSYHELTEAELLEQLS 1531


RESULT 6
PKSL_BACSU
ID   PKSL_BACSU     STANDARD;      PRT;  4427 AA.
AC   Q05470;
DT   01-FEB-1995 (Rel. 31, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Putative polyketide synthase pksL (PKS).
GN   PKSL OR PKSX OR PKSA OR OUTG.
OS   Bacillus subtilis.
OC   Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX   NCBI_TaxID=1423;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168 / PB1424;
RX   MEDLINE=93345824; PubMed=8344529;
RA   Scotti C., Piatti M., Cuzzoni A., Perani P., Tognoni A., Grandi G.,
RA   Galizzi A., Albertini A.M.;
RT   "A Bacillus subtilis large ORF coding for a polypeptide highly
RT   similar to polyketide synthases.";
RL   Gene 130:65-71(1993).
```

```
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168;
RX   MEDLINE=98044033; PubMed=9384377;
RA   Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA   Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
RA   Borriss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA   Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA   Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
RA   Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,
RA   Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA   Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA   Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
RA   Guiseppi G., Guy B.J., Haga K., Haiech J., Harwood C.R., Henaut A.,
RA   Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA   Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
RA   Kobayashi Y., Koetter P., Koningstein G., Krogh S., Kumano M.,
RA   Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA   Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,
RA   Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA   Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA   Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA   Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA   Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadaie Y.,
RA   Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA   Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,
RA   Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA   Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA   Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA   Viari A., Wambutt R., Wedler E., Wedler H., Weitzenegger T.,
RA   Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA   Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.;
RT   "The complete genome sequence of the Gram-positive bacterium Bacillus
RT   subtilis.";
RL   Nature 390:249-256(1997).
RN   [3]
RP   SEQUENCE OF 3619-4427 FROM N.A.
RC   STRAIN=168 / PB1424;
RA   Grandi G.;
RL   Submitted (JUL-1994) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: POTENTIALLY INVOLVED IN SOME INTERMEDIATE STEPS FOR
CC       THE SYNTHESIS OF A POLYKETIDE MOLECULE WHICH MAY BE INVOLVED IN
CC       SECONDARY METABOLISM.
CC   -!- COFACTOR: CONTAINS 5 COVALENTLY BOUND PHOSPHOPANTETHEINES
CC       (POTENTIAL).
CC   -!- SIMILARITY: CONTAINS 5 ACYL CARRIER DOMAINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z14098; CAA78479.1; -.
DR   EMBL; U11039; AAA85145.1; -.
DR   EMBL; Z99113; CAB13602.1; -.
DR   EMBL; Z35133; CAA84504.1; -.
DR   PIR; S25021; S25021.
DR   SubtiList; BG10698; pksL.
DR   InterPro; IPR002198; ADH_short.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR003880; Ppantne_attach.
DR   Pfam; PF00106; adh_short; 1.
DR   Pfam; PF00109; ketoacyl-synt; 4.
DR   Pfam; PF00550; pp-binding; 5.
DR   Pfam; PF02801; ketoacyl-synt_C; 4.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; 5.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 1.
DR   PROSITE; PS50075; ACP_DOMAIN; 5.
KW   Transferase; Acyltransferase; Antibiotic biosynthesis; NADP;
KW   Phosphopantetheine; Multifunctional enzyme; Repeat; Complete proteome.
```

```
FT   DOMAIN      211    280       ACYL CARRIER (ACP) 1.
FT   DOMAIN      382    759       BETA-KETOACYL SYNTHASE 1.
FT   DOMAIN      937   1115       DEHYDRATASE.
FT   DOMAIN     1409   1602       BETA-KETOACYL REDUCTASE 1.
FT   DOMAIN     1687   1759       ACYL CARRIER (ACP) 2.
FT   DOMAIN     1876   2253       BETA-KETOACYL SYNTHASE 2.
FT   DOMAIN     2491   2560       ACYL CARRIER (ACP) 3.
FT   DOMAIN     2632   2701       ACYL CARRIER (ACP) 4.
FT   DOMAIN     2823   3182       BETA-KETOACYL SYNTHASE 3.
FT   DOMAIN     3575   3776       BETA-KETOACYL REDUCTASE 2.
FT   DOMAIN     3854   3923       ACYL CARRIER (ACP) 5.
FT   DOMAIN     4019   4373       BETA-KETOACYL SYNTHASE 4.
FT   BINDING     243    243       PHOSPHOPANTETHEINE (POTENTIAL).
FT   BINDING    1723   1723       PHOSPHOPANTETHEINE (POTENTIAL).
FT   BINDING    2523   2523       PHOSPHOPANTETHEINE (POTENTIAL).
FT   BINDING    2664   2664       PHOSPHOPANTETHEINE (POTENTIAL).
FT   BINDING    3886   3886       PHOSPHOPANTETHEINE (POTENTIAL).
SQ   SEQUENCE   4427 AA;  493398 MW;  9612521E561AB9F2 CRC64;
```

```
  Query Match             9.8%;  Score 2042;  DB 1;  Length 4427;
  Best Local Similarity   22.4%;  Pred. No. 5.8e-85;
  Matches  949;  Conservative  585;  Mismatches 1520;  Indels 1190;  Gaps  161;

Qy        408 FAHLGFDSVMGVELRNRLSKATGLRLPVTLIFDHTTPAAVAARL------RTAALGHLDE 461
              :  :| :|   :|:   :|   |  |  ||:|:||| | ::| |         :| | | :
Db        236 YYQMGLNSSGLLEVVETISDKIGESLSPTLLFEHTTIAELSAFLAEEYAEHFSAAGSLGQ 295

Qy        462 -DTAPVPDSPSGH----GGTAA---ADDPIAIIGMACRFPGGVRSPKDLWELAASGGDAI 513
               : | || : |     |   |    |||||:| |:|    :  : |      | | :
Db        296 NERARVSDSINDHKTVEGSRPAPIEAAGDIAIIGLAGRYPKAA-NIHEFWNNLKEGKDCV 354

Qy        514 GPFPTDRGWPTEQRHAQDPTQP-GTFYPQGGGFLHDAAHFDAGFFGISPREALAMDPQQR 572
                | | |  : :  :  | | |    : |||:  |   ||  || |:|||  ||||:|
Db        355 SEIPESR-W--DWQRLEGITSPSGKDISKWGGFIDDPDCFDPQFFRITPREAETMDPQER 411

Qy        573 LLLETSWEAFERAGIDPLSV-------RGSRTGVFAGALSFDYGPRMDTASSEGAADVEG 625
              | ||| ||  | ||  | ::       :    ||||| :  ||     |      |  |
Db        412 LFLETCWETIEDAGYTPKTLAKPKGRNKRQHVGVFAGVMHKDY-----TLVGAEEASAEN 466

Qy        626 HILTGTTGSVLSGRIAYSFGLEGPAITVDTGCSASLVTLHLACQSLRSGECTLALAGGVS 685
                  : :: |::|     ||:: ||| ||:||  :||| :|:| ||| :||||||:
Db        467 VFPLSLNYAQIANRVSYFCNFHGPSMAVDTVCSSSLTAVHLALESIRHGECDVALAGGVN 526

Qy        686 VMSTLGMFIEFSRQRGLSVDGRCKAYSAAADGTGWGEGVGMLLVERLSDAVRLGHRVLAV 745
              :      :: :       | || |: :    ||    ||:| :|:: |  |   | |: ||
Db        527 LSLHPNKYMTYGVWDMFSTDGHCRTFGKDGDGYVPAEGIGAVLLKPLRQAEEDGDRIYAV 586

Qy        746 VRGSAVNQDGASNGLTAPNGPAQERVIRQALANAGLSVADVDVVEGHGTGTTLGDPIEAQ 805
              ::|||||  |  :|:: |:  :|  :|   |   |:    :  || |||||:|||||| |
Db        587 IKGSAVNHVGTVSGISVPSPVSQADLIETCLEKTGIDPRTISYVEAHGTGTSLGDPIEIQ 646

Qy        806 ALLATYGQRAGDRPLW-LGSLKSNIGHTMAAAGVGGVIKMVMALREGVLPRTLHVDKPSP 864
               |:  : |   ||    :|||:||||||  :|||: |: |: : |    | :|| :: :|
Db        647 GLVKAFRQYTQDRQFCSIGSVKSNIGHAESAAGISGLSKVALQLHHQKLVPSLHSEELNP 706

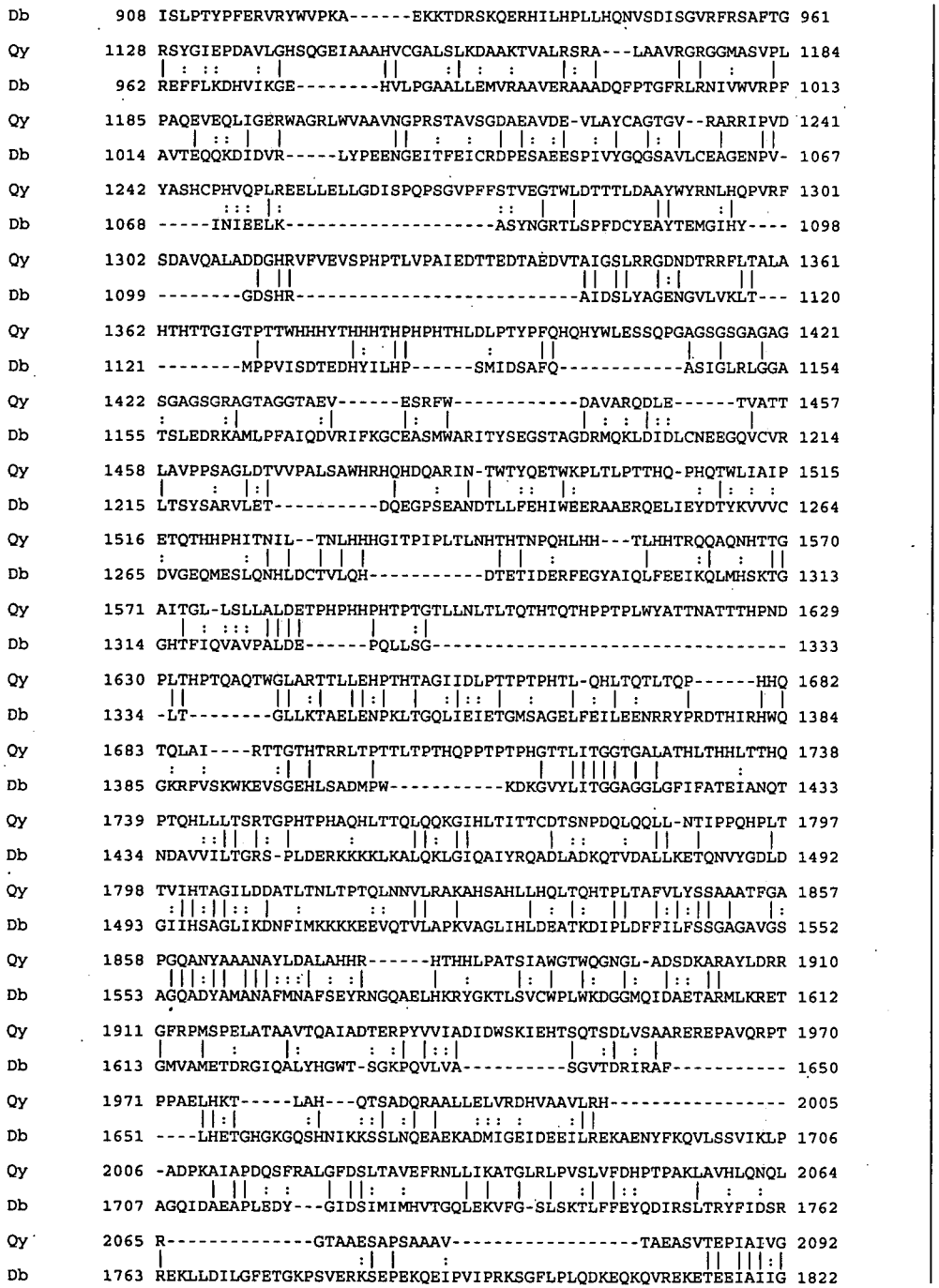
Qy        865 QVD--------------WSAGAVRLLTEAVPWPGDAAGRLRRAGVSSFGIGGTNAHVILE 910
               ||              |   ::   | ||:|      ||||:|||| |:|||:|||
Db        707 YVDFEKSPFYVQHETETWKQPVIKENGEDVPYP-------RRAGISSFGATGSNAHIILE 759

Qy        911 EAPAAGGCVAGGGVLEGAPGLAISVAESVAAPVAVSAPVAESVPVPVPVPVPVPVSARSE 970
              |                       : ::    |::|:           :  :|:|||::
Db        760 E---------------------HIPQAAEQDVSLSS--------DSDISAVIPLSARNQ 789

Qy        971 AGLRAQAEALRQYVAVRPDVSLADVGAGLACGRAVLEHRAVVLAADREELVQGLGALAAG 1030
               ||  |: |  :  :   : :|:   |  || :| |   ||:  :||  |  |   |
Db        790 ERLRVYAKRLLDF--LHDGIQIRDLAYTLQVGREPMEERVSFLASGIQELSDQLKAFIEG 847

Qy       1031 EP------DRRVTTGHAPGGDRGGVVFVFPGQ-----GGQWA-GMGV--RLL--ASSPVF 1074
                    ||: | |     |  :   :        ||| | ||  :||  | |
Db        848 RKAIQHCWKGRVSRGSEPSRPAESVHKLLEQRKLDQIAEQWANGSGVDWKLLYEGSKPKR 907

Qy       1075 ARRMQACEEALAPWVDWSVVDILRRDAGDAVWERADVVQPVLF-------SVMVSLAALW 1127
                    | :  ||  :          |   :   :: |:|         |    |
```


Db 908 ISLPTYPFERVRYWVPKA------EKKTDRSKQERHILHPLLHQNVSDISGVRFRSAFTG 961
Qy 1128 RSYGIEPDAVLGHSQGEIAAAHVCGALSLKDAAKTVALRSRA---LAAVRGRGGMASVPL 1184
Db 962 REFFLKDHVIKGE--------HVLPGAALLEMVRAAVERAAADQFPTGFRLRNIVWVRPF 1013
Qy 1185 PAQEVEQLIGERWAGRLWVAAVNGPRSTAVSGDAEAVDE-VLAYCAGTGV--RARRIPVD 1241
Db 1014 AVTEQQKDIDVR-----LYPEENGEITFEICRDPESAEESPIVYGQGSAVLCEAGENPV- 1067
Qy 1242 YASHCPHVQPLREELLELLGDISPQPSGVPFFSTVEGTWLDTTTLDAAYWYRNLHQPVRF 1301
Db 1068 -----INIEELK--------------------ASYNGRTLSPFDCYEAYTEMGIHY---- 1098
Qy 1302 SDAVQALADDGHRVFVEVSPHPTLVPAIEDTTEDTAEDVTAIGSLRRGDNDTRRFLTALA 1361
Db 1099 --------GDSHR---------------------------AIDSLYAGENGVLVKLT--- 1120
Qy 1362 HTHTTGIGTPTTWHHHYTHHHTHPHPHTHLDLPTYPFQHQHYWLESSQPGAGSGSGAGAG 1421
Db 1121 --------MPPVISDTEDHYILHP------SMIDSAFQ------------ASIGLRLGGA 1154
Qy 1422 SGAGSGRAGTAGGTAEV------ESRFW------------DAVARQDLE------TVATT 1457
Db 1155 TSLEDRKAMLPFAIQDVRIFKGCEASMWARITYSEGSTAGDRMQKLDIDLCNEEGQVCVR 1214
Qy 1458 LAVPPSAGLDTVVPALSAWHRHQHDQARIN-TWTYQETWKPLTLPTTHQ-PHQTWLIAIP 1515
Db 1215 LTSYSARVLET----------DQEGPSEANDTLLPEHIWEERAAERQELIEYDTYKVVVC 1264
Qy 1516 ETQTHHPHITNIL--TNLHHHGITPIPLTLNHTHTNPQHLHH---TLHHTRQQAQNHTTG 1570
Db 1265 DVGEQMESLQNHLDCTVLQH----------DTETIDERFEGYAIQLFEEIKQLMHSKTG 1313
Qy 1571 AITGL-LSLLALDETPHPHHPHTPTGTLLNLTLTQTHTQTHPPTPLWYATTNATTTHPND 1629
Db 1314 GHTFIQVAVPALDE------PQLLSG---------- 1333
Qy 1630 PLTHPTQAQTWGLARTTLLEHPTHTAGIIDLPTTPTPHTL-QHLTQTLTQP------HHQ 1682
Db 1334 -LT--------GLLKTAELENPKLTGQLIEIETGMSAGELFEILEENRRYPRDTHIRHWQ 1384
Qy 1683 TQLAI----RTTGTHTRRLTPTTLTPTHQPPTPTPHGTTLITGGTGALATHLTHHLTTHQ 1738
Db 1385 GKRFVSKWKEVSGEHLSADMPW----------KDKGVYLITGGAGGLGFIFATEIANQT 1433
Qy 1739 PTQHLLLTSRTGPHTPHAQHLTTQLQQKGIHLTITTCDTSNPDQLQQLL-NTIPPQHPLT 1797
Db 1434 NDAVVILTGRS-PLDERKKKKLKALQKLGIQAIYRQADLADKQTVDALLKETQNVYGDLD 1492
Qy 1798 TVIHTAGILDDATLTNLTPTQLNNVLRAKAHSAHLLHQLTQHTPLTAFVLYSSAAATFGA 1857
Db 1493 GIIHSAGLIKDNFIMKKKKEEVQTVLAPKVAGLIHLDEATKDIPLDFFILFSSGAGAVGS 1552
Qy 1858 PGQANYAAANAYLDALAHHR------HTHHLPATSIAWGTWQGNGL-ADSDKARAYLDRR 1910
Db 1553 AGQADYAMANAFMNAFSEYRNGQAELHKRYGKTLSVCWPLWKDGGMQIDAETARMLKRET 1612
Qy 1911 GFRPMSPELATAAVTQAIADTERPYVVIADIDWSKIEHTSQTSDLVSAARAREPAVQRPT 1970
Db 1613 GMVAMETDRGIQALYHGWT-SGKPQVLVA----------SGVTDRIRAF---------- 1650
Qy 1971 PPAELHKT-----LAH---QTSADQRAALLELVRDHVAAVLRH---------------- 2005
Db 1651 ----LHETGHGKGQSHNIKKSSLNQEAEKADMIGEIDEEILREKAENYFKQVLSSVIKLP 1706
Qy 2006 -ADPKAIAPDQSFRALGFDSLTAVEFRNLLIKATGLRLPVSLVFDHPTPAKLAVHLQNQL 2064
Db 1707 AGQIDAEAPLEDY---GIDSIMIMHVTGQLEKVFG-SLSKTLFFEYQDIRSLTRYFIDSR 1762
Qy 2065 R--------------GTAAESAPSAAAV----------------TAEASVTEPIAIVG 2092
Db 1763 REKLLDILGFETGKPSVERKSEPEKQEIPVIPRKSGFLPLQDKEQKQVREKETEEIAIIG 1822
Qy 2093 MACRFPGGVTSADDFWDLISSEQDAIGGFPTDRGWDLDTLYDPDPDHPGTCYTRNGGFLY 2152
Db 1823 ISGRYP-QADNIDELWEKLRDGRDCITEIPADR-WDHSLYYDEDKDKPGKTYSKWGGFMK 1880
Qy 2153 DAGHFDAEFFGISPREALAMDPQQRLLLETAWETIEHAGINPHTLH-------GTPTGVF 2205
Db 1881 DVDKFDPQFFHISPREAKLMDPQERLFLQCVYETMEDAGYTREHLGRKRDAELGGSVGVY 1940
Qy 2206 TGTNGQDYALRVHNAGQSTDG--FALTGTAGSVISGRISYTFGFEGPAVSVDTACSSSLV 2263
Db 1941 VGVMYEEYQL--YGAQEQVRGRSLALTGNPSS-IANRVSYYFDFHGPSIALDTMCSSSLT 1997
Qy 2264 ALHLACQALRAGECSMALAGGVTVMSSPGAFVEFSRQRGLAADGHCKAFSAAADGTGWGE 2323
Db 1998 AIHLACQSLQRGECEAAFAGGVNVSIHPNKYLMLGQNKFMSSKGRCESFGQGGDGYVPGE 2057
Qy 2324 GVGMLLVERLSDAHRNGHRVLAVVRGSAVNQDGASNGLTAPNGPSQQRVIRQALANAGLS 2383
Db 2058 GVGAVLLKPLSKAVEDGDHIYGIIKGTAINHGGKTNGYSVPNPNAQADVIKKAFVEAKVD 2117
Qy 2384 AGDVDAVEAHGTGTTLGDPIEAQALLATYGQDRAGEGPLWLGSVKSNVGHTQAAAGVAGV 2443
Db 2118 PRTVSYIEAHGTGTSLGDPIEITGLSKVFTQETDDKQFCAIGSAKSNIGHCESAAGIAGV 2177
Qy 2444 IKMVMALRHGLLPRTLHVDEPSPHVDWSAGAVQLLTETVPWP-------GGEGRLRR-AG 2495
Db 2178 TKVLLQMKYRHVAPSLHSNVLNPNIDFLNSPFKVQQELEEWKRPIISVNGKDIELPRIAG 2237
Qy 2496 VSSFGVSGTNAHVILEEAPADDVPGGPPAGEGDAGSDDEAAAGSPGVWPWLVSAKSQPAL 2555
Db 2238 VSSFGAGGVNAHILIEEYAPEPVEERLPARK----------QPAV--IVLSAKNEERL 2283
Qy 2556 RAQAQ-ALHA-HLTDHPGLDLADVGYTLAHARAVFDHRATLIAADRDTFLQALQALAAGE 2613
Db 2284 QKRPERLLHAIREQTYVEADLHRIAYTLQVGREAMKERLAFVAETMQELEEKLYECISGT 2343
Qy 2614 PHPAVIH-------SSAPGGTGTGEAAGKT--AFICSGQ----------G 2644
Db 2344 ENREYVYRGQVKSNKEAIAAFAADEDMSKTIEAWLQKGKYAKVLDLWVRGLRIDWSTLYQ 2403
Qy 2645 TQRPGMAHGLYHTHPVFAAALNDICTHLDPHLDH--------PLLPLLTQNDNDNEDAAA 2696
Db 2404 DQKPRRI-----SLPAYPFARDRYWIDVNAKAEEKRTEEPFAPVQPVIPKPSVDREAS-- 2456
Qy 2697 LLQQTRYAQPALFAFQVALHRLLTDGYHITPHYYAGHSLGEITAAHLAGILTLTDATTLI 2756
Db 2457 -------GKPA----NITLQPLMTN----------QDRLERVPSDTETETIT 2487
Qy 2757 TQRATLMQTMPPGTMTTLHTTPHHITHHLTAHENDLAIAAINTPTSLVISGTPHTVQHIT 2816
Db 2488 AE--ALCDELTAGLAEVLYMDQNEI-------DPDEAFIDIGMDS---ITG----LEWIK 2531
Qy 2817 TLCQQQG-----IKTKTLPTNHAFHSPHTNPILNQLHQHTQTLTYHP-PHTPLITANTPP 2870
Db 2532 AINKQYGTSLNVTKVYDYPTTRDFAVYLAHELSTQAGEKKQTETYTPIRQKTVVPAAKPA 2591
Qy 2871 DQLLTP--HYW--TQQARNTVDYA----------TTTQTLHQH---GVT---------- 2902
Db 2592 NISLQPLEHHQPVQEEAEETIQYAAAEISASRQYTVAIETLHENLRESIADVLYMEPYEV 2651
Qy 2903 ----TYIELGPDNTLTTLTHHNLPNPP-TTTLTLTHPHHHPQTHLLTNLAKTTTTWHPHH 2957
Db 2652 DIDEAFIDIGMD-SITGLEWIKAVNKQYGTSFTVTRVYDYPTIRDFAEMLKSELGTHLDR 2710
Qy 2958 YTHH-------DNQPHTHTH---LDLPTYPFQH-----HH----------YWL 2985
Db 2711 KIEHTDSFEAAQQKPAASSHPKPAERPLQPVQHPIKKEHEKKTVPVLQDRPEDAIAIVGM 2770
Qy 2986 ESTQPGAGNV----------SAAGLDPTE 3004
Db 2771 SGRYPGARNVREYWDNLVHARNAIRDIPTSRWDVDKYYDPVLNKKGKVYCKSMGMLDDIE 2830

```
Qy       3005 H--PLL-----------------------------------------GATLELATDGGAL 3021
              |  ||                                         |  | : ::   :
Db       2831 HFDPLFFNIPPSEAELMDPQHRIFLQEGYKAPEDAGYNARTLNEKKCGVYLGIMSNEYGV 2890

Qy       3022 LAGRLSLRSHPWLADHAVGGTVLLSGAT---FLEL--------ALHAGTYVGCDRVDELT 3070
              :  | |          :| | :  :: |    || |             : : ||     :
Db       2891 MLNRQS-------RANATGNSFAIAAARIPYFLNLKGPAIPIDTACSSSLVGTHLARQAL 2943

Qy       3071 LHAPLVVPVDGGVSV-----------QVGVAAADGEGRRLVSVYARGGSACGGGGASGGV 3119
              ::  : : : ||||:           : |: : ||: :             |    |
Db       2944 INKEIDMALVGGVSLYLTPESYMSMCEAGMLSPDGQCK-----------AFDNG------ 2986

Qy       3120 WTCHASGVLVEAAAGGVVVDGLAGVWPPRGAVAVDVDGVRDRLAGAGCVLGPVFSGLRAV 3179
                  |:| :    || :|:  |           | :  || :   | ::|   ||:
Db       2987 ----ANGFVPGEGAGALVLKRLK----------DAEADRDHI--YGIIIG---SGINQD 3026

Qy       3180 WRDGG----DLLAEVCLPEEAWGDAAGFGLHPALLDGVVQPLSVLLPGGTGFGEGAGFGE 3235
               :  |       ::: |  : :    :|:||       : :| :   |||  :    |:
Db       3027 GKTNGITAPSAKSQMDLERDIY---ETYGIHP-------ESISYVEMHGTGTKQ----GD 3072

Qy       3236 GVRVPAV---------------WGGVSLH------RAGVTGV----------------- 3256
              : : |:               | |  :     . ||| ||
Db       3073 PIELEALSTVFQEKTDKKQFCAIGSVKSNIGHTSAAAGVAGVQKVLLCMNHKTLVPTLNF 3132

Qy       3257 ------------------------------RVRVSAVGRGGGREAVSVVVGDEAGVPVA 3285
                                           |  ||: |  |    : :    |   |
Db       3133 TTPNEHFEFEHSPLYVNTELKPWETADGKPRRACVSSFGYSGTNAHIVI----EEYQPEK 3188

Qy       3286 SVDRL-----------------ELR--------------PVDMGQLRAVSVSAGR----- 3309
               |||                 :|:               :|:  : | ::  ||
Db       3189 RNDRLTKQHRSALFVLSAKKEKQLKAYAEAMKDFVTSNEDIDLEDM-AYTLQTGREAMDY 3247

Qy       3310 ----------------------RGSLYAVQ----------------------WAEVG 3322
                                    |::|                        | |
Db       3248 RMAFLADSREMLIKALDDYLAEMPNASIFAAHVKTKKSEIKLFETDHDAKALLQTWIEKK 3307

Qy       3323 PVPVCGQAWA------WHEDVGE----------------------SGGGP---------- 3344
              :   : |       |::  ||                       | |
Db       3308 RLEKVAELWVKGLQIDWNKLYGEYTPRRISLPAYPFAEEYYWLPTQEGEPETIATAMPQF 3367

Qy       3345 --VPGVVVLR-----CP-------------------------------------------DA 3356
                :|    ||     ||                                          |
Db       3368 ELMPKRCFLRKQWQPCPIEPAEMTNQTVAILANEETMALAEELSAYFSTYRIFDSQELDR 3427

Qy       3357 GAGGGGGGGGGGGVGEVVGGVLGVV---------QGWLG-----LERFAGSR----LVVV 3398
              :                | | | : ::          ||:      :|:   |:    :: |
Db       3428 VSAAD---------YEHVAGAIDLIGCGTSHEHSMGWINWLQKLIEQGRASKHHLTVLGV 3478

Qy       3399 TRGAVVAGPEDGPVDVVGASVWGLVRSAQAEHPDRFVLLDLDTDTGTDLDTGAGAGWGVD 3458
              |:| : |   :| | : |||  || |  |:|:           |    |::  |
Db       3479 TKG-LEAYANEG-VLLSGASRAGLYRMLQSEYSHL-------TSRHADMECEA------- 3522

Qy       3459 GGRVAAVVACGEPQLAVRGERLLAARLKRLESSGDVPAQRSGDTRARRSDVPAQRSGGVP 3518
                        :||     ||:|        :  :|  :      ||  :: ||:
Db       3523 ----------SHEELA----RLIAVEYYAKSTESEVCYRNGQRYRAYLTEQPAE---AAL 3565

Qy       3519 ARRSVDVSGREVLPWLSGGSVLVTGGTGVLGAAVARHLAGVCGVRDLLLVSRRGPDAPG- 3577
              : : |  |  :||          |:|||| ||   |||     ||: |:|: |   : |
Db       3566 SHKQVSFSTDKVL--------LITGGTRGLGLLCARHFVKTYGVKRLVLIGRE--ELPPR 3615

Qy       3578 -----------AEGLRA--ELAALGAEVRIVACDVGERREVVRLLEGVPAGC-PLTGVVH 3623
                        || ::|  ||  :||:|:::: |: :|  | : |: :       : ||:|
Db       3616 DQWNSVKISSLAEKIKAVQELEDMGAQVQVLSLDLTDRVAVEQSLKTIHETMGAIGGVIH 3675

Qy       3624 AAGVLDDATIASL--TPERLGTVFAAKVDAALLLDELTRGMELSAFVLFSSAAGILG--S 3679
              ||::     | :  : | :| |   ||:    | :|  :   |: | |||| :  :   :
Db       3676 CAGMVNKQNPAFIRKSLEEIGQVLEPKVEGLQTLFDLLQDEPLAFFTLFSSVSAAIPALA 3735

Qy       3680 AGQGNYAAANAALDALAYRRRAAGLPGVSLAWGLWEEASGMTGHLAGTDHRRIIRSGLHP 3739
              ||| :|| ||| :|  |   :    | ||: |  |:|    || |     : : ::||
Db       3736 AGQADYAMANAFMDYFAEAHQDK-CPIVSIQWPNWKE----TG-LGEVRSKALEQTGLIS 3789

Qy       3740 MSTPDALALFDAALALDR--PVLLPA--DLRPAPPLPPLLQDLLPATRRRTTRTTTTGGA 3795
```

```
              ::  : | | |  |: ||   |:||| |     |   |:|  ||
Db       3790 LTNDEGLQLLDQILS-DRQYAVVLPAVPDTNVWKP-DKLMQPSLP--------------- 3832

Qy       3796 DNGAQLHARLAGQTHEQQHT-TLLALVRSHIATVLG---HTTPDTIPPDRAFRDLGFDSL 3851
                    : |    :| || |  |       : |:          :    |  |:: | ||:
Db       3833 -----VEALSHPETKEQTSTRNLFPETVDWLVTLFSDELKIAAEDFETDEPFQEYGIDSI 3887

Qy       3852 TAVELRNRLSRTTGLRLPTTLAFDHPNPTTLTHHLHTQ------LQPQPDNAVAPVLAEL 3905
                 :|  ::::     :  :: |::|   :  | | ::      |||       |:  :
Db       3888 ILAQLVQQMNQQLNGDIDPSILFEYPTIESFAHWLISKYDISRVLQPSVPEKQTPLKPQS 3947

Qy       3906 DKLESALSALDKTDSASERVTLRLKSLMLRWNAPQHPTAESADD 3949
                    :  :       : |:  |  : : :   : : | ||: ::
Db       3948 AMKQKLVPEQRPQQISHEKTALLAEDIAIIGLSCRFPGAETLEE 3991
```

```
RESULT 7
PPSA_MYCTU
ID   PPSA_MYCTU     STANDARD;      PRT;  1876 AA.
AC   Q10977;
DT   01-OCT-1996 (Rel. 34, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Phenolpthiocerol synthesis polyketide synthase ppsA.
GN   PPSA OR RV2931 OR MT3000 OR MTCY338.20.
OS   Mycobacterium tuberculosis.
OC   Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC   Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1773;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H37Rv;
RX   MEDLINE=98295987; PubMed=9634230;
RA   Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA   Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekaia F.,
RA   Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA   Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA   Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA   Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA   Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA   Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT   "Deciphering the biology of Mycobacterium tuberculosis from the
RT   complete genome sequence.";
RL   Nature 393:537-544(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CDC 1551 / Oshkosh;
RA   Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA   Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA   Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA   Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA   Bishai W.;
RT   "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT   laboratory strains.";
RL   Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: POLYKETIDE SYNTHASE INVOLVED IN PHENOLPTHIOCEROL
CC       SYNTHESIS.
CC   -!- COFACTOR: CONTAINS 2 COVALENTLY BOUND PHOSPHOPANTETHEINES (By
CC       similarity).
CC   -!- SIMILARITY: CONTAINS 2 ACYL CARRIER DOMAINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z74697; CAA98988.1; -.
DR   EMBL; AE007122; AAK47328.1; -.
DR   TIGR; MT3000; -.
DR   TubercuList; Rv2931; -.
```

```
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR003880; Ppantne_attach.
DR   Pfam; PF00109; ketoacyl-synt; 1.
DR   Pfam; PF00550; pp-binding; 2.
DR   Pfam; PF00698; Acyl_transf; 1.
DR   Pfam; PF02801; ketoacyl-synt_C; 1.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; 1.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 1.
DR   PROSITE; PS50075; ACP_DOMAIN; 2.
KW   Multifunctional enzyme; Oxidoreductase; Transferase; NADP; Repeat;
KW   Phosphopantetheine; Complete proteome.
FT   DOMAIN        7     80       ACYL CARRIER (ACP) 1.
FT   DOMAIN     1764   1833       ACYL CARRIER (ACP) 2.
FT   NP_BIND    1498   1503       NADP (POTENTIAL).
FT   BINDING      62     62       PHOSPHOPANTETHEINE (POTENTIAL).
FT   ACT_SITE    273    273       BETA-KETOACYL SYNTHASE (BY SIMILARITY).
FT   ACT_SITE    720    720       MALONYLTRANSFERASE (BY SIMILARITY).
FT   BINDING    1796   1796       PHOSPHOPANTETHEINE (POTENTIAL).
FT   CONFLICT    624    624       D -> E (IN REF. 2).
FT   CONFLICT    877    877       R -> H (IN REF. 2).
FT   CONFLICT   1323   1323       G -> S (IN REF. 2).
SQ   SEQUENCE   1876 AA;  198834 MW;  D9783DBD48792110 CRC64;

  Query Match             9.7%;  Score 2017;  DB 1;  Length 1876;
  Best Local Similarity   30.0%;  Pred. No. 2.7e-84;
  Matches  591;  Conservative  243;  Mismatches  720;  Indels  418;  Gaps   51;

Qy        398 TASDAGPSASFAHLGFDSVMGVELRNRLSKATGLRLPVTLIFDHTTPAAVAARLRTAALG 457
Db         26 TPDEVDPDLSLADLGVSSRDAVVLSGELSELLGRTVSPIDFWEHPTINALAAYLAA---- 81

Qy        458 HLDEDTAPVPDSPSG--HGGTAAADDPIAIIGMACRFPGGVRSPKDLWELAASGGDAIGP 515
Db         82 -----PEPSPDSDAAVKRGARNSLDEPIAVVGMGCRFPGGISCPEALWDFLCERRSSISQ 136

Qy        516 FPTDR-----GWPTEQRHAQDPTQPGTFYPQGGGFLHDAAHFDAGFFGISPREALAMDPQ 570
Db        137 VPPQRWQPFEGGPPEVAAALART------TRWGSFLPDIDAFDAEFFEISPSEADKMDPQ 190

Qy        571 QRLLLETSWEAFERAGIDPLSVRGSRTGVFAGALSFDYGPRMDTASSEGAADVEGHILTG 630
Db        191 QRLLLEVAWEALEHAGIPPGTLRRSATGVFAGACLSEYG----AMASADLSQVDGWSNSG 246

Qy        631 TTGSVLSGRIAYSFGLEGPAITVDTGCSASLVTLHLACQSLRSGECTLALAGGVSVMSTL 690
Db        247 GAMSIIANRLSYFLDLRGPSVAVDTACSSSLVAIHLACQSLRTQDCHLAIAAGVNLLLSP 306

Qy        691 GMFIEFSRQRGLSVDGRCKAYSAAADGTGWGEGVGMLLVERLSDAVRLGHRVLAVVRGSA 750
Db        307 AVFRGFDQVGALSPTGQCRAFDATADGFVRGEGAGVVVLKRLTDAQRDGDRVLAVICGSA 366

Qy        751 VNQDGASNGLTAPNGPAQERVIRQALANAGLSVADVDVVEGHGTGTTLGDPIEAQALLAT 810
Db        367 VNQDGRSNGLMAPNPAAQMAVLRAAYTNAGMQPSEVDYVEAHGTGTLLGDPIEARALGTV 426

Qy        811 YGQ-RAGDRPLWLGSLKSNIGHTMAAAGVGGVIKMVMALREGVLPRTLHVDKPSPQVDWS 869
Db        427 LGRGRPEDSPLLIGSVKTNLGHTEAAAGIAGFIKTVLAVQHGQIPPNQHFETANPHIPFT 486

Qy        870 AGAVRLLTEAVPWPGDAAGRLRRAGVSSFGIGGTNAHVILEEAPAAGGCVAGGGVLEGAP 929
Db        487 DLRMKVVDTQTEWP--ATGHPRRAGVSSFGFGGTNAHVVIEQ----------GQEVRPAP 534

Qy        930 GLAISVAESVAAPVAVSAPVAESVPVPVPVPVPVPVSARSEAGLRAQAEALRQYV-AVRP 988
Db        535 GQGLSPAVST----------------------LVVAGKTMQRVSATAGMLADWMEGPGA 571

Qy        989 DVSLADVGAGLACGRAVLEHRAVVLAADREELVQGLGALAAGEPDRRVTTGHAPG----- 1043
Db        572 DVALADVAHTLNHHRSRQPKFGTVVARDRTQAIAGLRALAAGQ--------HAPGVVNPA 623

Qy       1044 -GDRG-GVVFVFPGQGGQWAGMGVRLLASSPVFARRMQACEEALAPWVDWSVVDILRRDA 1101
```

```
Db        624 DGSPGPGTVFVYSGRGSQWAGMGRQLLADEPAFAAAVAELEPVFVEQAGFSLHDVL---A 680

Qy       1102 GDAVWERADVVQPVLFSVMVSLAALWRSYGIEPDAVLGHSQGEIAAAHVCGALSLKDAAK 1161
Db        681 NGEELVGIEQIQLGLIGMQLALTELWCSYGVRPDLVIGHSMGEVAAAVVAGALTPAEGLR 740

Qy       1162 TVALRSRALAAVRGRGGMASVPLPAQEVEQLIGERWAGRLWVAAVNGPRSTAVSGDAEAV 1221
Db        741 VTATRSRLMAPLSGQGGMALLELDAPTTEALIAD--FPQVTLGIYNSPRQTVIAGPTEQI 798

Qy       1222 DEVLAYCAGTGVRARRIPVDYASHCPHVQPLREELLELLGDISPQPSGVPFFSTVEGTWL 1281
Db        799 DELIARVRAQNRFASRVNIEVAPHNPAMDALQPAMRSELADLTPRTPTIGIIST---TYA 855

Qy       1282 DTTT---LDAAYWYRNLHQPVRFSDAVQAL---ADDGHRVFVEVSPHPTLVPAIEDTTED 1335
Db        856 DLHTQPVFDAEHWATNMRNPVRFQQAIASAGSGADGAYHTFIEISAHPLLTQAIIDTLHS 915

Qy       1336 TAEDV--TAIGSLRRGDNDTRRFLTALAHTHTTGIGTPTTWHHHYTHHHTHPHPHTHLDL 1393
Db        916 AQPGARYTSLGTLQRDTDDVVTFRTNLNKAHTI--------HPPHTPHPPEPHP----PI 963

Qy       1394 PTYPFQHQHYWLESSQPGAGSGSGAGAGSGAGSGRAGTAGGTAEVESRFW------DAVA 1447
Db        964 PTTPWQHTRHWITTKYPAGSVGSAPRAGTLLGQ-HTTVATVSASPPSHLWQARLAPDAKP 1022

Qy       1448 RQDLETVATTLAVPPSAGLDTVVPALSAWHRHQHDQARINTWTYQETWKPLTL------- 1500
Db        1023 YQGGHRFHQVEVVPASVVLHTILSAATELGYSALSEVRFEQPIFADRPRLIQVVADNRAI 1082

Qy       1501 -----PTTHQPHQTWLIAI----------------------PETQTHHPHITNILTNL 1531
Db        1083 SLASSPAAGTPSDRWTRHVTAQLSSSPSDSASSLNEHHRANGQPPERAHRDLIPDLAELL 1142

Qy       1532 HHHGITPIPLT----------------------LNHTHTNP-----QHL-------HHTL 1557
Db        1143 AMRGIDGLPFSWTVASWTQHSSNLTVAIDLPEALPEGSTGPLLDAAVHLAALSDVADSRL 1202

Qy       1558 HHTRQQAQNHTTGAITGLLSLLALDETPHPH----------HPHTPTGTL--------- 1597
Db        1203 YVPASIEQISLGDVVTGPRSSVTLNRTAHDDDGITVDVTVAAHGEVPSLSMRSLRYRALD 1262

Qy       1598 --LNLTLTQTHTQTHP----------PTPLWYATTNATTTHPN-DPLTHP--------- 1634
Db        1263 FGLDVGRAQPPASTGPVEAYCDATNFVHTIDWQPQTVPDATHPGAEQVTHPGPVAIIGDD 1322

Qy       1635 ------------------------------------------------------------ 1634
Db        1323 GAALCETLEGAGYQPAVMSDGVSQARYVVYVADSDPAGADETDVDFAVRICTEITGLVRT 1382

Qy       1635 ----------------------------TQAQTWGLARTTLLEHPTHTAGIIDLPTT--- 1663
Db        1383 LAERDADKPAALWILTRGVHESVAPSALRQSFLWGLAGVIAAEHPELWGGLVDLAINDDL 1442

Qy       1664 ----PTPHTLQHLTQTLTQPHHQTQLAIRTTG-THTRRLTPTTLTPTHQPPTPTPHGTTL 1718
Db        1443 GEFGPA------LAELLAKP--SKSILVRRDGVVLAPALAPVRGEPARKSLQCRPDAAYL 1494

Qy       1719 ITGGTGALATHLTHHLTTHQPTQHLLLTSRTGPHTP----HAQHLTTQ----------LQ 1764
Db        1495 ITGGLGALGLLMADWL-ADRGAHRLVLTGRT-PLPPRRDWQLDTLDTELRRRIDAIRALE 1552

Qy       1765 QKGIHLTITTCDTSNPDQLQQLL--NTIPPQHPLTTVIHTAGILDDATLTNLTPTQLNNV 1822
Db        1553 MRGVTVEAVAADVGCREDVQALLAARDRDGAAPIRGIIHAAGITNDQLVTSMTGDAVRQV 1612

Qy       1823 LRAKAHSAHLLHQLTQHTPLTAFVLYSSAAATFGAPGQANYAAANAYLDALAHHRHTHHL 1882
Db        1613 MWPKIGGSQVLHDAFPPGSVDFFYLTASAAGIFGIPGQGSYAAANSYLDALARARRQQGC 1672

Qy       1883 PATSIAWGTWQGNGL-ADSDKARAYLDRRGFRPMSPELATAAVTQAIADTERPYVVIADI 1941
```

```
Db       1673 HTMSLDWVAWRGLGLAADAQLVSEELARMGSRDITPSEAFTA------------------ 1714

Qy       1942 DWSKIEHTSQTSDLVSAARAREREPAVQRPTP-------------PAELHKTLAHQTSADQR 1988
                :  ::      |: |       |  | |              ||     :|   : : |
Db       1715 -WEFVD----GYDVAQAV-----VVPMPAPAGADGSGANAYLLPARNWSVMA---ATEVR 1761

Qy       1989 AALLELVRDHVAAVLRHADPKAIAPDQSFRALGFDSLTAVEFRNLLIKATGLRLPVSLVF 2048
              : | : :|  :|| ||    | :  |: |  || :|| |:  |    :  |: |  :::|
Db       1762 SELEQGLRRIIAAELR-VPEKELDTDRPFAELGLNSLMAMAIRREAEQFVGIELSATMLF 1820

Qy       2049 DHPTPAKLAVHL-----------QNQLRGTAA-------------ESAPSAA 2076
                :|||   || :|           ||:   ::             ||||  |
Db       1821 NHPTVKSLASYLAKRVAPHDVSQDNQISALSSSAGSVLDSLFDRIESAPPEA 1872
```

```
RESULT 8
MCAS_MYCBO
ID   MCAS_MYCBO     STANDARD;      PRT;  2110 AA.
AC   Q02251;
DT   01-JUN-1994 (Rel. 29, Created)
DT   01-JUN-1994 (Rel. 29, Last sequence update)
DT   15-JUL-1998 (Rel. 36, Last annotation update)
DE   Mycocerosic acid synthase.
GN   MAS.
OS   Mycobacterium bovis.
OC   Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC   Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1765;
RN   [1]
RP   SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX   MEDLINE=92406887; PubMed=1527058;
RA   Mathur M., Kolattukudy P.E.;
RT   "Molecular cloning and sequencing of the gene for mycocerosic acid
RT   synthase, a novel fatty acid elongating multifunctional enzyme, from
RT   Mycobacterium tuberculosis var. bovis Bacillus Calmette-Guerin.";
RL   J. Biol. Chem. 267:19388-19395(1992).
CC   -!- FUNCTION: CATALYZES THE ELONGATION OF N-FATTY ACYL-COA WITH
CC       METHYLAMALONYL-COA (NOT MALONYL-COA) AS THE ELONGATING AGENT TO
CC       FORM MYCOCEROSYL LIPIDS.
CC   -!- COFACTOR: CONTAINS 1 COVALENTLY BOUND PHOSPHOPANTETHEINE.
CC   -!- SUBUNIT: HOMODIMER WHOSE MONOMERS PROBABLY HAVE A HEAD TO TAIL
CC       ARRANGEMENT.
CC   -!- SUBCELLULAR LOCATION: Membrane-associated.
CC   -!- SIMILARITY: CONTAINS 1 ACYL CARRIER DOMAIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M95808; AAA25369.1; -.
DR   PIR; B44110; B44110.
DR   HSSP; P73283; 1E5M.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR002085; Adh_zn_family.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR003880; Ppantne_attach.
DR   Pfam; PF00107; adh_zinc; 1.
DR   Pfam; PF00109; ketoacyl-synt; 1.
DR   Pfam; PF00550; pp-binding; 1.
DR   Pfam; PF00698; Acyl_transf; 1.
DR   Pfam; PF02801; ketoacyl-synt_C; 1.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; 1.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 1.
DR   PROSITE; PS50075; ACP_DOMAIN; 1.
KW   Fatty acid biosynthesis; Multifunctional enzyme; Phosphopantetheine;
KW   Transferase; Hydrolase; Oxidoreductase; Ligase; NADP; Membrane.
FT   DOMAIN        1    430       BETA-KETOACYL SYNTHASE.
FT   DOMAIN      533    852       ACYL TRANSFERASE.
FT   DOMAIN        ?      ?       ENOYL REDUCTASE.
FT   DOMAIN        ?      ?       BETA-KETOACYL REDUCTASE.
FT   DOMAIN     2026   2096       ACYL CARRIER (ACP).
FT   ACT_SITE    177    177       BETA-KETOACYL SYNTHASE (BY SIMILARITY).
FT   ACT_SITE    623    623       ACYL TRANSFERASES (BY SIMILARITY).
FT   NP_BIND    1561   1578       NADP (ER).
FT   NP_BIND    1765   1780       NADP (KR).
FT   BINDING    2059   2059       PHOSPHOPANTETHEINE (BY SIMILARITY).
SQ   SEQUENCE   2110 AA;  225577 MW;  CC658215D7155300 CRC64;
```

```
  Query Match             8.8%;  Score 1832.5;  DB 1;  Length 2110;
  Best Local Similarity   27.5%;  Pred. No. 7.9e-76;
  Matches  603;  Conservative  222;  Mismatches  678;  Indels  693;  Gaps   64;

Qy        482 PIAIIGMACRFPGGVRSPKDLWELAASGGDAIGPFPTDRGWPTEQRHAQDPTQPGTFYPQ 541
              |:|:||| || |||: ||  |||     | | :   | || |  :  :  :|  ||    :
Db          7 PVAVIGMGCRLPGGINSPDKLWESLLRGDDLVTEIPPDR-WDADDYYDPEPGVPGRSVSR 65

Qy        542 GGGFLHDAAHFDAGFFGISPREALAMDPQQRLLLETSWEAFERAGIDPLSVRGSRTGVFA 601
              |||| | | ||| ||||| ||| ::|||||||||||||| | ||:|| |: || | ||
Db         66 WGGFLDDVAGFDAEFFGISEREATSIDPQQRLLLETSWEAIEHAGLDPASLAGSSTAVFT 125

Qy        602 GALSFDYGPRMDTASSEGAADVEGHILTGTTGSVLSGRIAYSFGLEGPAITVDTGCSASL 661
              |    ||   : | :: | |    :::||   || |||||:: || ||| :| || || |
Db        126 GLTHEDY--LVLTTTAGGLA--SPYVVTGLNNSVASGRIAHTLGLHGPAMTFDTACSSGL 181

Qy        662 VTLHLACQSLRSGECTLALAGGVSVMSTLGMFIEFSRQRGLSVDGRCKAYSAAADGTGWG 721
              : :||||:||  ||  |||||| :|:      :  | |  ||  ||| :: | |||
Db        182 MAVHLACRSLHDGERDLALAGGCAVLLEPHACVAASAQGMLSSTGRCHSFDADADGFVRS 241

Qy        722 EGVGMLLVERLSDAVRLGHRVLAVVRGSAVNQDGASNGLTAPNGPAQERVIRQALANAGL 781
              ||  |:|::|| ||:| |:|: |||||:| |||| :  || |:  ||  | | ||| ||:
Db        242 EGCAMVLLKRLPDALRDGNRIFAVVRGTATNQDGRTETLTMPSEDAQVAVYRAALAAAGV 301

Qy        782 SVADVDVVEGHGTGTTLGDPIEAQALLATYGQRAGDRPLWLGSLKSNIGHTMAAAGVGGV 841
                | ||| ||||| :|||||| ::|   ||  ||  |  ||| ||| |:|| |:||  |:
Db        302 QPETVGVVEAHGTGTPIGDPIEYRSLARVYG--AG-TPCALGSAKSNMGHSTASAGTVGL 358

Qy        842 IKMVMALREGVLPRTLHVDK-PSPQVDWSAGAVRLLTEAV-PWPGDAAGRLRRAGVSSFG 899
              || :::|| ||:|  || :: |    |    |    : :|| |||       :|  |||||
Db        359 IKAILSLRHGVVPPLLHFNRLPDELSDVETGL--FVPQAVTPWPNGNDHTPKRVAVSSFG 416

Qy        900 IGGTNAHVILEEAPAAGGCVAGGGVLEGAPGLAISVAESVAAPVAVSAPVAESVPVPVPV 959
              : ||| | |:|||||                              |||  || |    |
Db        417 MSGTNVHAIVEEAPA----------------------------EASAP--ESSPGDAEV 445

Qy        960 -PVPVPVSARSEAGLRAQAEALRQYVAVRPD-VSLADVGAGLACGRAVLEHRAVVLAADR 1017
               |    :|: |   || |  |  :|    | |: :|:   || ||     |  |:||:
Db        446 GPRLFMLSSTSSDALRQTARQLATWVEEHQDCVAASDLAYTLARGRRHRPVRTAVVAANL 505

Qy       1018 EELVQGLGALAAGEPDRRVTTGHAPGGDRGGVVFVFPGQGGQWAGMGVRLLASSPVFARR 1077
              |||:||  :| |:        ||   ||| | |:|| ||| ||| || :|||| ||||
Db        506 PELVEGLREVADGDALYDAAVGH---GDR-GPVWVFSGQGSQWAAMGTQLLASEPVFAAT 561

Qy       1078 MQACEEALAPWVDWSVVDILRRDAGDAVWERADVVQPVLFSVMVSLAA-LWRSYGIEPDA 1136
              :   |  :|    :|| : :             | ||| :|:| |:||| : ::||: | |
Db        562 IAKLEPVIAAESGFSVTEAI---TAQQTVTGIDKVQPAVFAVQVALAATMEQTYGVRPGA 618

Qy       1137 VLGHSQGEIAAAHVCGALSLKDAAKTVALRSRALAAVRGRGGMASVPLPAQEVEQLIGER 1196
              |:||| || ||| | |||||:|||: :  ||: :  : | |:| || |||::|   :  |
Db        619 VVGHSMGESAAAVVAGALSLEDAARVICRRSKLMTRIAGAGAMGSVELPAKQVNSELMAR 678

Qy       1197 WAGRLWVAAVNGPRSTAVSGDAEAVDEVLAYCAGTGVRARRIPVDYASHCPHVQPLREEL 1256
                : |: |  |:|| : | :: | :::|       | || : || ||| | | |: ::|
Db        679 GIDDVVVSVVASPQSTVIGGTSDTVRDLIARWEQRDVMAREVAVDVASHSPQVDPILDDL 738

Qy       1257 LELLGDISPQPSGVPFFSTVEGTWLDTTTLDAAYWYRNLHQPVRFSDAVQALADDGHRVF 1316
                | ||:|   ||::|      :    | |||  ||   |:|: |||| :||:|||
Db        739 AAALADIAPMTPKVPYYSATLFDPREQPVCDGAYWVDNLRNTVQFAAAVQAAMEDGYRVF 798

Qy       1317 VEVSPHPTLVPAIEDTTEDTAEDVTAIGSLRRGD---NDTRRFLTALAHTHTTG------ 1367
                :||||  |  |:| |       | |:  :||      :  |  || |   | |
Db        799 ADVSPHPLLTHAVEQTGRSLDMSVAALAGMRREQPLPHGLRGLLTEL---HRAGAALDYS 855
```

```
Qy       1368 --------IGTP-TTWHHHYTHHH------------------THPHPHTHLDLPTYPFQ 1399
                      :  |   |   ||                    ||   :|: |   | :
Db        856 ALYPAGRLVDAPLPAW----THARLFIDDDGQEQRAQGACTITVHPLLGSHVRLTEEPER 911

Qy       1400 HQHY---------WLESSQ-------PGAG---------------------------- 1413
                |           ||   |       |||
Db        912 HVWQGDVGTSVLSWLSDHQVHNVAALPGAAYCEMALAAAAAEVFGEAAEVRDITFEQMLLL 971

Qy       1414 ------------------------------------------------------------ 1413

Db        972 DEQTPIDAVASIDAPGVVNFTVETNRDGETTRHATAALRAAEDDCPPPGYDITALLQAHP 1031

Qy       1414 ----------SGSGAGAGSGAGSG----------------------RAGTAGGTAEVES 1440
                      | :  |   ||  |                        |:|::    |
Db       1032 HAVNGTAMRESFAERGVTLGAAFGGLTTAHTPRREPRRCWPRSRCPRRSGSSRAPTESTR 1091

Qy       1441 RFWDAVARQDLETVATTLAVPPSAGLDTVVPA-LSAWHR-----HQHDQA---------- 1484
                | |  |: :     :      |:  :    || | |: :     |  ||
Db       1092 RCWTLVSSRSARASSRYGHWWPAVAVGCAQPACLRAYPQCPLLLHAVDQGLQRRDPRWEA 1151

Qy       1485 ------------------RINTWTYQET--------------WKPLTLPTTHQPHQ-TWL 1511
                              |: | | :                |:   ||       :||
Db       1152 DLDVLDEHGTVLLAVRGLRMGTGTSERDERDRLVSERLLTLGWQQRALPEVGDGEAGSWL 1211

Qy       1512 IAIPETQTHHPH-ITNILTN-LHHHG-----ITPIPLTLNHTHTNPQ------------- 1551
                :          |  : : ||: |  ||         :  ::   |  | |
Db       1212 LIDTSNAVDTPDMLASTLTDALKSHGPQGTECASLSWSVQDTPPNDQAGLEKLGSQLRGR 1271

Qy       1552 ----------------------HLHHTLHHTRQ----------------QAQ---NHT 1568
                                    : | :  ||:                  |||    |
Db       1272 DGVVIVYGPRVGDPDEHSLLAGREQVRHLVRITRELAEFEGELPRLFVVTRQAQIVKPHD 1331

Qy       1569 TG-----AITGLLSLLALDETPHPHHPHTPTGTLLNLTL--TQTHTQTH----------- 1610
                :|        ||  || :  : ||         :|  ||     ||
Db       1332 SGERANLEQAGLRGLLRVISSEHP---------MLRTTLIDVDEHTDVERVAQQLLSGSE 1382

Qy       1611 ----------------PPTPLWYATTNATTTHPND-----PLTHPTQAQT---------- 1639
                             |:|| :          |:       :  |   ||
Db       1383 EDETAWRNGDWYVARLTPSPLGHEERRTAVLDPDHDGMRVQVRRPGDLQTLEFVASDRVP 1442

Qy       1640 ------------------------------------------------------------ 1639

Db       1443 PGPGQIEVAVSMSSINFADVLIAFGRFPIIDDREPQLGMDFVGVVTAVGEGVTGHQVGDR 1502

Qy       1640 ---------W----------------GL----ARTTLLEHPTHTAGIIDL---------- 1660
                         |               ||    | |    | |   |: ||
Db       1503 VGGFSEGGCWRTFLTCDANLAVTLPPGLTDEQAITAATAHATAWYGLNDLAQIKAGDKVL 1562

Qy       1661 ----------------------PTTPTP----------H----------------- 1667
                                      |  |          |
Db       1563 IHSATGGVGQAAISIARAKGAEIFATAGNPAKRAMLRDMGVEHVYDSRSVEFAEQIRRDT 1622

Qy       1668 -------TLQHLT------------------------------------------------ 1673
                     |  ||
Db       1623 DGYGVDIVLNSLTGAAQRAGLELLPFGGRFVEIGKADVYGNTRLGLFPFRRGLTFYYLDL 1682

Qy       1674 --QTLTQPHHQTQL---------------------------AIR--TTGTHTRRLT--- 1698
                  ::|||    :|                         |||  :   || :|
Db       1683 ALMSVTQPDRVRELLATVFKLTADGVLTAPQCTHYPLAEAADAIRAMSNAEHTGKLVLDV 1742

Qy       1699 ------PTTLTPTHQPPTPTPHGTTLITGGTGALATHLTHHLTTHQPTQHLLLTSRTGPH 1752
                     :||  | |      |: :|||| | |         |        ::||:|: |:
Db       1743 PRSGRRSVAVTP-EQRPLYRRDGSYIITGGLGGLGLFFASKLAA-AGCGRIVLTARSQPN 1800

Qy       1753 TPHAQHLTTQLQQKGIHLTITTCDTSNPDQLQQLLNTIPPQH-PLTTVIHTAGILDDATL 1811
                | |:     |:  |  : :   : : ||   :|::       ||  |:|:| :::||||
Db       1801 -PKARQTIEGLRAAGADIVVECGNIAEPDTADRLVSAATATGLPLRGVLHSAAVVEDATL 1859

Qy       1812 TNLTPTQLNNVLRAKAHSAHLLHQLTQHTPLTAFVLYSSAAATFGAPGQANYAAANAYLD 1871
                ||:|   ::     |   :  ||: |   ||  | |:|| ||  |:|||  |||||:::|
Db       1860 TNITDELIDRDWSPKVFGSWNLHRATLGQPLDWFCLFSSGAALLGSPGQGAYAAANSWVD 1919

Qy       1872 ALAHHRHTHHLPATSIAWGTWQGNGLADSDKARAYLDRRGFRPMSPELATAAVTQAIADT 1931
```

```
                  ||      || ::|||| |   | |       :|   |   ::||    |  : :
Db       1920 VFAHWPRAQGLPVSAIAWGAWGEVGRA------TFLAEGGEIMITPEEGAYAF-ETLVRH 1972

Qy       1932 ERP---YVVIADIDWSKIEHTSQTSDLVSAARAREP-------AVQRPTPPAELHKTLAH 1981
                :|    |: |     |         :|||     | |              ||   |::    |
Db       1973 DRAYSGYIPILGAPW--------LADLV----RRSPWGEMFASTGQRSRGPSKFRMELLS 2020

Qy       1982 QTSADQRAALLELVRDHVAAVLRHADPKAIAPDQSFRALGFDSLTAVEFRNLLIKATGLR 2041
                    :    |  |: :  : :||    : |  |:||   | |||  :| |  :   ||:|
Db       2021 LPQDEWAGRLRRLLVEQASVILR----RTIDADRSFIEYGLDSLGMLEMRTHVETETGIR 2076

Qy       2042 LPVSLVFDHPTPAKLAVHLQNQLRGTAAESAPSAAA 2077
                |   ::  : |   || :| : |   | | | : ||
Db       2077 LTPKVIATNKTARALAQYLADTL---AEEQAAAPAA 2109
```

RESULT 9
PKSK_BACSU
ID PKSK_BACSU STANDARD; PRT; 4447 AA.
AC P40803;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Putative polyketide synthase pksK (PKS).
GN PKSK.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168 / PB1424;
RX MEDLINE=95219083; PubMed=7704258;
RA Albertini A.M., Caramori T., Scoffone F., Scotti C., Galizzi A.;
RT "Sequence around the 159 degree region of the Bacillus subtilis
RT genome: the pksX locus spans 33.6 kb.";
RL Microbiology 141:299-309(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=98044033; PubMed=9384377;
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
RA Borriss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,
RA Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
RA Guiseppi G., Guy B.J., Haga K., Haiech J., Harwood C.R., Henaut A.,
RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
RA Kobayashi Y., Koetter P., Koningstein G., Krogh S., Kumano M.,
RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,
RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadaie Y.,
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,
RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA Viari A., Wambutt R., Wedler E., Wedler H., Weitzenegger T.,
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.;
RT "The complete genome sequence of the Gram-positive bacterium Bacillus
RT subtilis.";
RL Nature 390:249-256(1997).
CC -!- FUNCTION: POTENTIALLY INVOLVED IN SOME INTERMEDIATE STEPS FOR
CC THE SYNTHESIS OF A POLYKETIDE MOLECULE WHICH MAY BE INVOLVED IN

```
CC       SECONDARY METABOLISM.
CC    -!- COFACTOR: CONTAINS 5 COVALENTLY BOUND PHOSPHOPANTETHEINES
CC       (POTENTIAL).
CC    -!- SIMILARITY: BELONGS TO THE ATP-DEPENDENT AMP-BINDING ENZYME
CC       FAMILY.
CC    -!- SIMILARITY: CONTAINS 5 ACYL CARRIER DOMAINS.
CC    ---------------------------------------------------------------------
CC    This SWISS-PROT entry is copyright. It is produced through a collaboration
CC    between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC    the European Bioinformatics Institute. There are no restrictions on its
CC    use by non-profit institutions as long as its content is in no way
CC    modified and this statement is not removed. Usage by and for commercial
CC    entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC    or send an email to license@isb-sib.ch).
CC    ---------------------------------------------------------------------
DR    EMBL; U11039; AAA85144.1; -.
DR    EMBL; Z99112; CAB13590.1; -.
DR    EMBL; Z99113; CAB13601.1; -.
DR    HSSP; P14687; 1AMU.
DR    SubtiList; BG10930; pksK.
DR    InterPro; IPR000873; AMP-bind.
DR    InterPro; IPR001242; Condensatn.
DR    InterPro; IPR000794; Ketoacyl-synt.
DR    InterPro; IPR003880; Ppantne_attach.
DR    Pfam; PF00109; ketoacyl-synt; 3.
DR    Pfam; PF00501; AMP-binding; 1.
DR    Pfam; PF00550; pp-binding; 5.
DR    Pfam; PF00668; Condensation; 1.
DR    Pfam; PF02801; ketoacyl-synt_C; 3.
DR    PRINTS; PR00154; AMPBINDING.
DR    PROSITE; PS00012; PHOSPHOPANTETHEINE; 3.
DR    PROSITE; PS00455; AMP_BINDING; 1.
DR    PROSITE; PS00606; B_KETOACYL_SYNTHASE; 2.
DR    PROSITE; PS50075; ACP_DOMAIN; 5.
KW    Transferase; Acyltransferase; Antibiotic biosynthesis; NADP;
KW    Phosphopantetheine; Multifunctional enzyme; Repeat; Ligase;
KW    Complete proteome.
FT    DOMAIN       1     68       ACYL CARRIER (ACP) 1.
FT    DOMAIN    1063   1130       ACYL CARRIER (ACP) 2.
FT    DOMAIN    2516   2589       ACYL CARRIER (ACP) 3.
FT    DOMAIN    2618   2687       ACYL CARRIER (ACP) 4.
FT    DOMAIN    3868   3937       ACYL CARRIER (ACP) 5.
FT    BINDING     31     31       PHOSPHOPANTETHEINE (POTENTIAL).
FT    BINDING   1093   1093       PHOSPHOPANTETHEINE (POTENTIAL).
FT    BINDING   2552   2552       PHOSPHOPANTETHEINE (POTENTIAL).
FT    BINDING   2650   2650       PHOSPHOPANTETHEINE (POTENTIAL).
FT    ACT_SITE  2915   2915       BETA-KETOACYL SYNTHASE (BY SIMILARITY).
FT    BINDING   3900   3900       PHOSPHOPANTETHEINE (POTENTIAL).
FT    ACT_SITE  4147   4147       BETA-KETOACYL SYNTHASE (BY SIMILARITY).
SQ    SEQUENCE   4447 AA;  496058 MW;  9C6DB4A6C0A9C057 CRC64;

  Query Match             8.7%;  Score 1809;  DB 1;  Length 4447;
  Best Local Similarity   20.9%;  Pred. No. 2.3e-74;
  Matches  857;  Conservative  514;  Mismatches 1301;  Indels 1424;  Gaps  136;

Qy        190 GPRSTVVSGAREAVADLVADLTAAQVRTRMIPVDVPAHSPLMYAIEERVVSGLLPITPRP 249
              ||  | :  | : :           |      :  | :  :| : :       : | |
Db        848 GPTETTIWSAVQRI----------NVECSHATIGRPIANTQIYITDSQ-------LAPVP 890

Qy        250 SRIPFHSSVTG-----GRLDTRELDAAYWYRNMSSTVRFEPAARLL--------LQQGPK 296
              : :|   : |     |    ||  : :  |      ||| ::|          | |
Db        891 AGVPGELCIAGDGVAKGYYKKEELTDSRFIDN-----PFEPGSKLYRTGDMARWLTGGRI 945

Qy        297 TFVEMSPHPVLTMG----LQELAPDLGDTTGTADTVIMG---------TLRRGQGTLD-- 341
               ::    : |   |    | ::   | :  |  : |::           | :    :|
Db        946 EYIGRIDNQVKIRGFRIELGDIESRLSEHPGILECVVVADMDNLAAYYTAKHANASLTAR 1005

Qy        342 ---HFLTSL---------------AQLRGHGETSATTVLSARLTALSPTQQQSL---LLD 380
                 ||: :                 | :|:    :: :  |:      |:|:    : |
Db       1006 ELRHFVKNALPAYMVPSYFIQLDHMPLTPNGKIDRNSLKNIDLSGEQLKQRQTSPKNIQD 1065

Qy        381 LVRAHTMAVLN------DDGNERTASDAGPSASFAHLGFDSVMGVELRNRLSKATGLRLP 434
                |       ||       |||             |  :| ||:: | : :|:
```

```
Db       1066 TVFTIWQEVLKTSDIEWDDG------------FFDVGGDSLLAVTVADRIKHELSCEFS 1112

Qy        435 VTLIFDHTTPAAVAARLRTAALG-----------HLDEDTAPVPDSPSGHGGTAAADDP 482
              || :|:::|    ::  :   :|           |::: :  : | | :     ||
Db       1113 VTDLFEYSTIKNISQYITEQRMGDASDHIPTDPAAHIEDQSTEMSDLPDYY------DDS 1166

Qy        483 IAIIGMACRFPGGVRSPKDLWELAASGGDAIGPFPTDR----GWPTEQRHAQDPTQPGTF 538
              :||||::| || | ::  : ||     | ::|  |  :     |    |      |
Db       1167 VAIIGISCEFP-GAKNHDEFWENLRDGKESIAFFNKEELQRFGISKEIAENAD------- 1218

Qy        539 YPQGGGFLHDAAHFDAGFFGISPREALAMDPQQRLLLETSWEAFERAGIDPLSVRGSRTG 598
              |      :     ||  || |||::|  |||| |:||  ||:| | ||    :  :|
Db       1219 YVPAKASIDGKDRFDPSFFQISPKDAEFMDPQLRMLLTHSWKAIEDAGYAARQI--PQTS 1276

Qy        599 VFAGALSFDYGPRMDTASSEGAADVEGHI--LTGTTGSVLSGRIAYSFGLEGPAITVDTG 656
              ||  | :  |   : : ::|     :|::  :   :|:: :  |::  || ||:  |
Db       1277 VFMSASNNSYRALLPSDTTESLETPDGYVSWVLAQSGTIPT-MISHKLGLRGPSYFVHAN 1335

Qy        657 CSASLVTLHLACQSLRSGECTLALAGGVSVMSTLGMFIEFSRQRGL--SVDGRCKAYSAA 714
              ||:||: || | :|| |||   || || :: :     | :  | ||  | ||  ||: |:
Db       1336 CSSSLIGLHSAYKSLLSGESDYALVGGATLHTESN--IGYVHQPGLNFSSDGHIKAFDAS 1393

Qy        715 ADGTGWGEGVGMLLVERLSDAVRLGHRVLAVVRGSAVNQDGASN-GLTAPNGPAQERVIR 773
              |||   |||| ::|::: :|||: |  : |::||  || |||   |  ||:   |  |::
Db       1394 ADGMIGGEGVAVVLLKKAADAVKDGDHIYALLRGIGVNNDGADKVGFYAPSVKGQADVVQ 1453

Qy        774 QALANAGLSVADVDVVEGHGTGTTLGDPIEAQALLATYGQRAGDRPL-WLGSLKSNIGHT 832
              | :    :    : || ||||| ||||||  ||   | |          :||:|:||||
Db       1454 QVMNQTKVQPESICYVEAHGTGTKLGDPIELAALTNVYRQYTNKTQFCGIGSVKTNIGHL 1513

Qy        833 MAAAGVGGVIKMVMALREGVLPRTLHVDKPSPQVDWSAGAVRLLTEAVPWPGDAAGRLRR 892
              |||: | ||:||:|     | :::  :|:|  | ::    :: :      :  :  |
Db       1514 DTAAGLAGCIKVVMSLYHQELAPSVNYKEPNPNTDLASSPFYVVDQKKTLSREI--KTHR 1571

Qy        893 AGVSSFGIGGTNAHVILEE-------APAAGGCVAGGGVLEGAPGLAISVAESVAAPVAV 945
              | :||||:|||| | | |:          | |:
Db       1572 AALSSFGLGGTNTHAIFEQFKRDSDKGKIDGTCI-------------------------- 1605

Qy        946 SAPVAESVPVPVPVPVPVPVSARSEAGLRAQAEALRQYVAVR--PDVSLADVGAGLACGR 1003
                              ||:||:::  |:  || :  |:  |   :  | |    |  ||
Db       1606 -----------------VPISAKNKERLQEYAEDILAYLERRGFENSQLPDFAYTLQVGR 1648

Qy       1004 AVLEHRAVVLAADREELVQGL----------GALAAGEPDRRVTTGHAPGGDRGGVVFV 1052
              :||| | :|    || | |                |  : | :  |      |  ::
Db       1649 EAMEHRVVFIADHVNELKQRLTDFINGNTAIEGCFQGSKHNAREVSWLTEDEDSAELIRK 1708

Qy       1053 FPGQG------GQWAGMGVRL---------------LASSPVFARRMQACEEALAPWVD- 1090
              : :|        |:  |  :                 | : |    |   ::   |
Db       1709 WMAKGKVNKLAEMWS-KGAHIDWMQLYKGERPNRMSLPTYPFAKERYWPSQDDRKPVAQI 1767

Qy       1091 -------WSVVDILRRD-------------AGDAVWERADVV--QPVLFSVM---VSLAA 1125
                    |:  :| ::               ||  : |  ||  :|||  |    :: ||
Db       1768 SGNQTGIGSIHPLLHQNTSDFSEQKFSSVFTGDEFFLRDHVVRGKPVLPGVAYLEMAYAA 1827

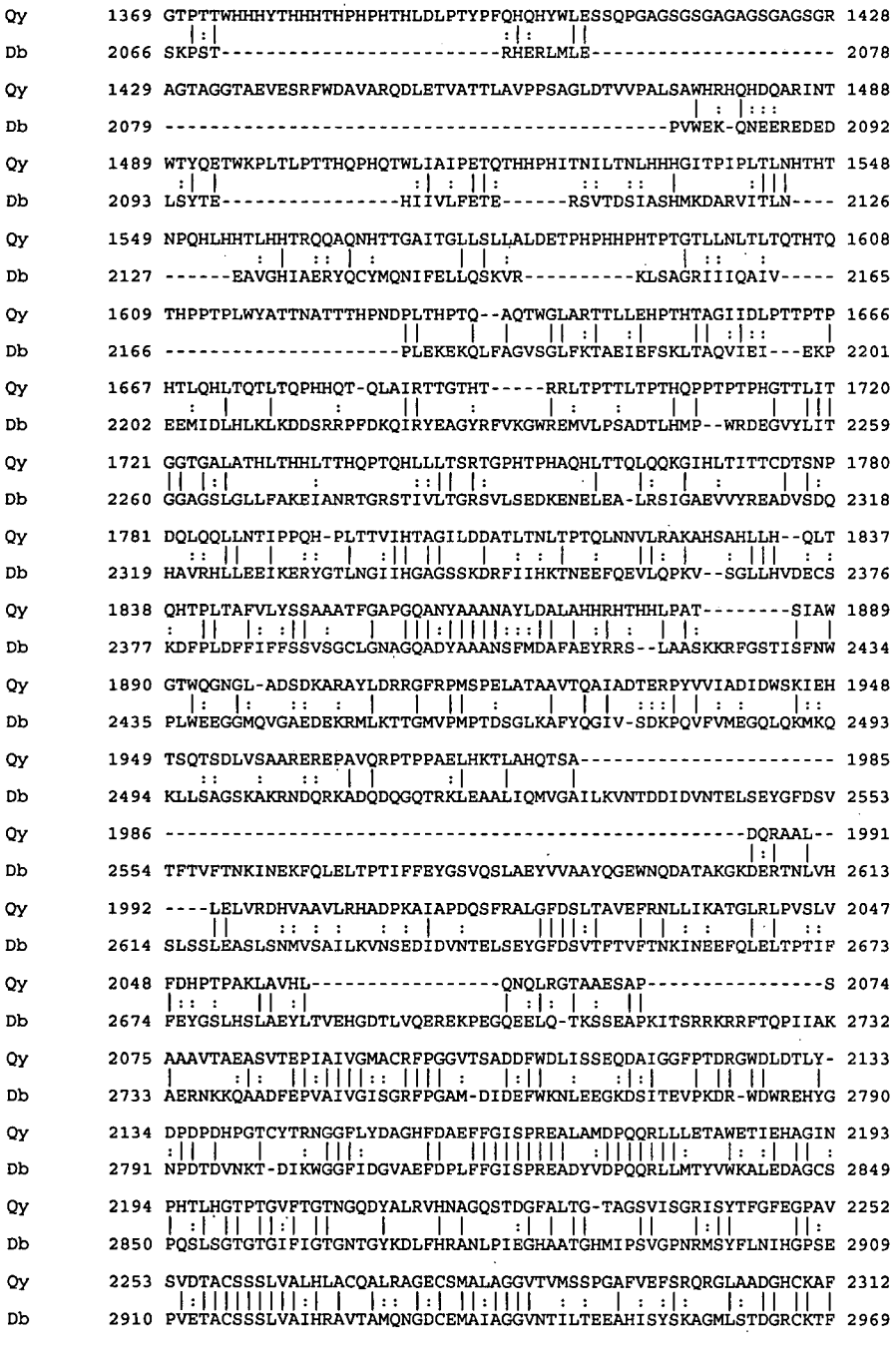
Qy       1126 LWRSYG----------------IEPDAVLGHS-----------------------QGEI 1145
              : :: |                ::|  |  ||                       |:
Db       1828 INQAAGSEIGQDVRIRLNHTVWVQPVVVDRHSAQVDISLFPEEDGKITFDIYSTQEDGDD 1887

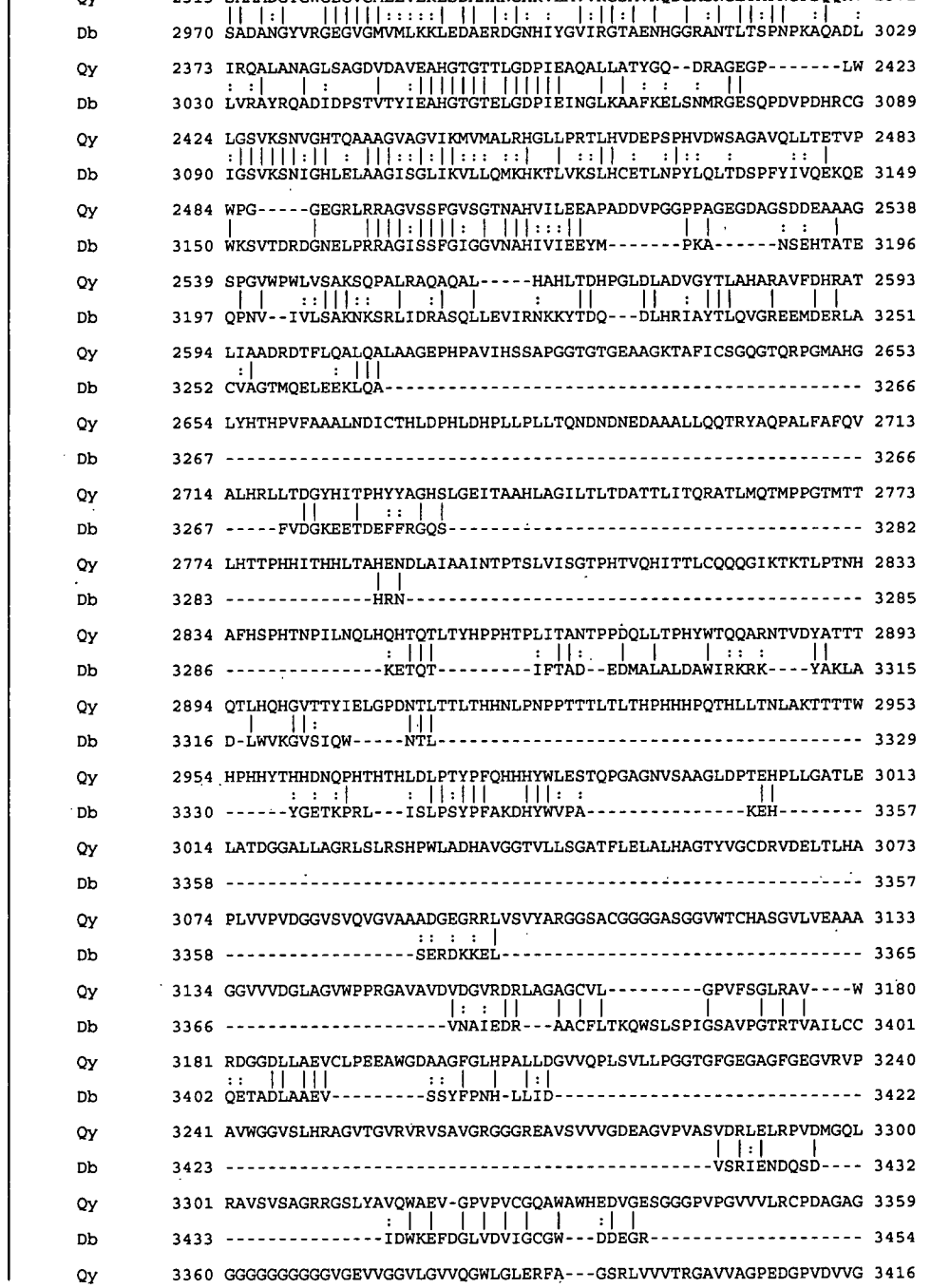
Qy       1146 AAAHVCGALSLKDAAKTVALRSRALAAVRGRGGMASVPLPAQEVEQLIGERWAGRLWVAA 1205
              |  |:  |  ||:|       ::    |:| |:    | |  |:    |  |    |
Db       1888 PVIHSQGSAELASAAETPVADLTEMSRRCGKGKMS----PDQFYEE---GRSRGMFHGPA 1940

Qy       1206 VNGPRSTAVSGDAEAVDEVLA------YCAGTGVRARRIPVDYASHCPHV--QPLREELL 1257
                | ::  :        ||||        :||  :        :  | | :     |:  :
Db       1941 FQGIKNVNIGN-----REVLAQLQLPEIVSGTNEQ-------FVLH-PSIMDSALQTATI 1987

Qy       1258 ELLGDISPQPSGVPF----FSTVEGTWLDTTTLDAAYW-YRNLHQPVRFSDAVQA----L 1308
              :: ::: |   :||      ::|          :: | |  |          ||     :
Db       1988 CIMQELTDQKLILPFALEELEVIKG-------CSSSMWAYARLSDSDHSGGVVQKADIDV 2040

Qy       1309 ADDGHRVFVEVSPHPTLVPAIEDTTEDTAEDVTAIGSLRRGDNDTRRFLTALAHTHTTGI 1368
              |:   | | :    |                          | |    ||
Db       2041 IDESGTVCVRIKGFST----------------------------RVLEGEVHT----- 2065
```


Qy 1369 GTPTTWHHHYTHHHTHPHPHTHLDLPTYPFQHQHYWLESSQPGAGSGSGAGAGSGAGSGR 1428
Db 2066 SKPST----------RHERLMLE---------- 2078
Qy 1429 AGTAGGTAEVESRFWDAVARQDLETVATTLAVPPSAGLDTVVPALSAWHRHQHDQARINT 1488
Db 2079 ----------PVWEK-QNEEREDED 2092
Qy 1489 WTYQETWKPLTLPTTHQPHQTWLIAIPETQTHHPHITNILTNLHHHGITPIPLTLNHTHT 1548
Db 2093 LSYTE----------HIIVLFETE------RSVTDSIASHMKDARVITLN---- 2126
Qy 1549 NPQHLHHTLHHTRQQAQNHTTGAITGLLSLLALDETPHPHHPHTPTGTLLNLTLTQTHTQ 1608
Db 2127 ------EAVGHIAERYQCYMQNIFELLQSKVR----------KLSAGRIIIQAIV----- 2165
Qy 1609 THPPTPLWYATTNATTTHPNDPLTHPTQ--AQTWGLARTTLLEHPTHTAGIIDLPTTPTP 1666
Db 2166 ----------PLEKEKQLFAGVSGLFKTAEIEFSKLTAQVIEI---EKP 2201
Qy 1667 HTLQHLTQTLTQPHHQT-QLAIRTTGTHT-----RRLTPTTLTPTHQPPTPTPHGTTLIT 1720
Db 2202 EEMIDLHLKLKDDSRRPFDKQIRYEAGYRFVKGWREMVLPSADTLHMP--WRDEGVYLIT 2259
Qy 1721 GGTGALATHLTHHLTTHQPTQHLLLTSRTGPHTPHAQHLTTQLQQKGIHLTITTCDTSNP 1780
Db 2260 GGAGSLGLLFAKEIANRTGRSTIVLTGRSVLSEDKENELEA-LRSIGAEVVYREADVSDQ 2318
Qy 1781 DQLQQLLNTIPPQH-PLTTVIHTAGILDDATLTNLTPTQLNNVLRAKAHSAHLLH--QLT 1837
Db 2319 HAVRHLLEEIKERYGTLNGIIHGAGSSKDRFIIHKTNEEFQEVLQPKV--SGLLHVDECS 2376
Qy 1838 QHTPLTAFVLYSSAAATFGAPGQANYAAANAYLDALAHHRHTHHLPAT--------SIAW 1889
Db 2377 KDFPLDPFIFFSSVSGCLGNAGQADYAAANSFMDAFAEYRRS--LAASKKRFGSTISFNW 2434
Qy 1890 GTWQGNGL-ADSDKARAYLDRRGFRPMSPELATAAVTQAIADTERPYVVIADIDWSKIEH 1948
Db 2435 PLWEEGGMQVGAEDEKRMLKTTGMVPMPTDSGLKAFYQGIV-SDKPQVFVMEGQLQKMKQ 2493
Qy 1949 TSQTSDLVSAARERERPAVQRPTPPAELHKTLAHQTSA---------- 1985
Db 2494 KLLSAGSKAKRNDQRKADQDQGQTRKLEAALIQMVGAILKVNTDDIDVNTELSEYGFDSV 2553
Qy 1986 ----------DQRAAL-- 1991
Db 2554 TFTVFTNKINEKFQLELTPTIFFEYGSVQSLAEYVVAAYQGEWNQDATAKGKDERTNLVH 2613
Qy 1992 ----LELVRDHVAAVLRHADPKAIAPDQSFRALGFDSLTAVEFRNLLIKATGLRLPVSLV 2047
Db 2614 SLSSLEASLSNMVSAILKVNSEDIDVNTELSEYGFDSVTFTVFTNKINEEFQLELTPTIF 2673
Qy 2048 FDHPTPAKLAVHL----------QNQLRGTAAESAP----------S 2074
Db 2674 PEYGSLHSLAEYLTVEHGDTLVQEREKPEGQEELQ-TKSSEAPKITSRRKRRFTQPIIAK 2732
Qy 2075 AAAVTAEASVTEPIAIVGMACRFPGGVTSADDFWDLISSEQDAIGGFPTDRGWDLDTLY- 2133
Db 2733 AERNKKQAADFEPVAIVGISGRFPGAM-DIDEFWKNLEEGKDSITEVPKDR-WDWREHYG 2790
Qy 2134 DPDPDHPGTCYTRNGGFLYDAGHFDAEFFGISPREALAMDPQQRLLLETAWETIEHAGIN 2193
Db 2791 NPDTDVNKT-DIKWGGFIDGVAEFDPLFFGISPREADYVDPQQRLLMTYVWKALEDAGCS 2849
Qy 2194 PHTLHGTPTGVFTGTNGQDYALRVHNAGQSTDGFALTG-TAGSVISGRISYTFGFEGPAV 2252
Db 2850 PQSLSGTGTGIFIGTGNTGYKDLFHRANLPIEGHAATGHMIPSVGPNRMSYFLNIHGPSE 2909
Qy 2253 SVDTACSSSLVALHLACQALRAGECSMALAGGVTVMSSPGAFVEFSRQRGLAADGHCKAF 2312
Db 2910 PVETACSSSLVAIHRAVTAMQNGDCEMAIAGGVNTILTEEAHISYSKAGMLSTDGRCKTF 2969


Qy 2313 SAAADGTGWGEGVGMLLVERLSDAHRNGHRVLAVVRGSAVNQDGASNGLTAPNGPSQQRV 2372
Db 2970 SADANGYVRGEGVGMVMLKKLEDAERDGNHIYGVIRGTAENHGGRANTLTSPNPKAQADL 3029
Qy 2373 IRQALANAGLSAGDVDAVEAHGTGTTLGDPIEAQALLATYGQ--DRAGEGP-------LW 2423
Db 3030 LVRAYRQADIDPSTVTYIEAHGTGTELGDPIEINGLKAAFKELSNMRGESQPDVPDHRCG 3089
Qy 2424 LGSVKSNVGHTQAAAGVAGVIKMVMALRHGLLPRTLHVDEPSPHVDWSAGAVQLLTETVP 2483
Db 3090 IGSVKSNIGHLELEAAGISGLIKVLLQMKHKTLVKSLHCETLNPYLQLTDSPFYIVQEKQE 3149
Qy 2484 WPG-----GEGRLRRAGVSSFGVSGTNAHVILEEAPADDVPGGPPAGEGDAGSDDEAAAG 2538
Db 3150 WKSVTDRDGNELPRRAGISSFGIGGVNAHIVIEEYM-------PKA------NSEHTATE 3196
Qy 2539 SPGVWPWLVSAKSQPALRAQAQAL-----HAHLTDHPGLDLADVGYTLAHARAVFDHRAT 2593
Db 3197 QPNV--IVLSAKNKSRLIDRASQLLEVIRNKKYTDQ---DLHRIAYTLQVGREEMDERLA 3251
Qy 2594 LIAADRDTFLQALQALAAGEPHPAVIHSSAPGGTGTGEAAGKTAFICSGQGTQRPGMAHG 2653
Db 3252 CVAGTMQELEEKLQA---------- 3266
Qy 2654 LYHTHPVFAAALNDICTHLDPHLDHPLLPLLTQNDNDNEDAAALLQQTRYAQPALFAFQV 2713
Db 3267 ---------- 3266
Qy 2714 ALHRLLTDGYHITPHYYAGHSLGEITAAHLAGILTLTDATTLITQRATLMQTMPPGTMTT 2773
Db 3267 -----FVDGKEETDEFFRGQS---------- 3282
Qy 2774 LHTTPHHITHHLTAHENDLAIAAINTPTSLVISGTPHTVQHITTLCQQQGIKTKTLPTNH 2833
Db 3283 ----------HRN---------- 3285
Qy 2834 AFHSPHTNPILNQLHQHTQTLTYHPPHTPLITANTPPDQLLTPHYWTQQARNTVDYATTT 2893
Db 3286 ----------KETQT---------IFTAD--EDMALALDAWIRKRK----YAKLA 3315
Qy 2894 QTLHQHGVTTYIELGPDNTLTTLTHHNLPNPPTTTLTLTHPHHHPQTHLLTNLAKTTTTW 2953
Db 3316 D-LWVKGVSIQW-----NTL---------- 3329
Qy 2954 HPHHYTHHDNQPHTHTHLDLPTYPFQHHHYWLESTQPGAGNVSAAGLDPTEHPLLGATLE 3013
Db 3330 ------YGETKPRL---ISLPSYPFAKDHYWVPA----------KEH-------- 3357
Qy 3014 LATDGGALLAGRLSLRSHPWLADHAVGGTVLLSGATFLELALHAGTYVGCDRVDELTLHA 3073
Db 3358 ---------- 3357
Qy 3074 PLVVPVDGGVSVQVGVAAADGEGRRLVSVYARGGSACGGGGASGGVWTCHASGVLVEAAA 3133
Db 3358 ----------SERDKKEL---------- 3365
Qy 3134 GGVVVDGLAGVWPPRGAVAVDVDGVRDRLAGAGCVL---------GPVFSGLRAV----W 3180
Db 3366 ----------VNAIEDR---AACFLTKQWSLSPIGSAVPGTRTVAILCC 3401
Qy 3181 RDGGDLLAEVCLPEEAWGDAAGFGLHPALLDGVVQPLSVLLPGGTGFGEGAGFGEGVRVP 3240
Db 3402 QETADLAAEV---------SSYFPNH-LLID---------- 3422
Qy 3241 AVWGGVSLHRAGVTGVRVRVSAVGRGGGREAVSVVVGDEAGVPVASVDRLELRPVDMGQL 3300
Db 3423 ----------VSRIENDQSD---- 3432
Qy 3301 RAVSVSAGRRGSLYAVQWAEV-GPVPVCGQAWAWHEDVGESGGGPVPGVVVLRCPDAGAG 3359
Db 3433 ----------IDWKEFDGLVDVIGCGW---DDEGR---------- 3454
Qy 3360 GGGGGGGGGGVGEVVGGVLGVVQGWLGLERFA---GSRLVVVTRGAVVAGPEDGPVDVVG 3416

```
                                |  ::    |  |    | ||: ||:|  :   ::  | : |
Db       3455 -----------------LDWIEWVQRLVEFGHKEGLRLLCVTKG--LESFQNTSVRMAG 3494

Qy       3417 ASVWGLVRSAQAEHPDRFVLLDLDTDTGTDLDTGAGAGWGVDGGRVAAVVA------CGE 3470
              ||  || |  | |:    :   :| :  |            |  |:| ::|       :
Db       3495 ASRAGLYRMLQCEY-SHLISRHMDAEEVT-----------DHRRLAKLIADEFYSDSYD 3541

Qy       3471 PQLAVRGERLLAARLKRLESSGDVPAQRSGDTRARRSDVPAQRSGGVPARRSVDVSGREV 3530
               ::  |      | ||    :|    | :                              |
Db       3542 AEVCYRDGLRYQAFLKAHPETGKATEQSA----------------------------V 3571

Qy       3531 LPWLSGGSVLVTGGTGVLGAAVARHLAGVCGVRDLLLVSR-----------RGPDAPGA 3578
               |      :|:||||  :|   ||| |   ||: |:|  |           :  :   |
Db       3572 FP--KDHVLLITGGTRGIGLLCARHFAECYGVKKLVLTGREQLPPREEWARFKTSNTSLA 3629

Qy       3579 EGLRA--ELAALGAEVRIVACDVGERREVVRLLEGVPAGC-PLTGVVHAAGVLDDATIAS 3635
              | ::|  || | | :| ::: : :  :| : |: :     |: ||:| ||: |  |:|
Db       3630 EKIQAVRELEAKGVQVEMLSLTLSDDAQVEQTLQHIKRTLGPIGGVIHCAGLTDMDTLAF 3689

Qy       3636 L--TPERLGTVFAAKVDAALLLDELTRGMELSAFVLFSSAAGILG--SAGQGNYAAANAA 3691
              :  | : :  |   ||      |        |  |||||| : |:   |||| :|| ||:
Db       3690 IRKTSDDIQRVLEPKVSGLTTLYRHVCNEPLQFFVLFSSVSAIIPELSAGQADYAMANSY 3749

Qy       3692 LD--ALAYRRRAAGLPGVSLAWGLWEEASGMTGHLAGTDHRRIIRSGLHPMSTPDALALF 3749
              :|  | |::: |   | :|: |  |:| :|| | :    :|    |||  ::  : |
Db       3750 MDYFAEAHQKHA---PIISVQWPNWKE-TGM-GEVTNQAYR---DSGLLSITNSEGLRFL 3801

Qy       3750 DAALALD-RPVLLPADLRPAPPLPPLLQDLLPATRRRTTRTTTTGGADNGAQLHARLAGQ 3808
              |  ::    ||:|||        | ||       :||       ||     | | :  |
Db       3802 DQIVSKKFGPVVLPAMANQTNWEPELLM------KRRKPH---EGGLQEAA-LQSPPARD 3851

Qy       3809 THEQQHTT----LLALVRSHIATVLG---HTTPDTIPPDRAFRDLGFDSLTAVELRNRLS 3861
                |    :    ||:  :| :  :         :    |  |:| | ||:   ::  |::
Db       3852 IEEADEVSKCDGLLSETQSWLIDLFTEELRIDREDFEIDGLFQDYGVDSIILAQVLQRIN 3911

Qy       3862 RTTGLRLPTTLAFDHP 3877
              |    |  :: :::|
Db       3912 RKLEAALDPSILYEYP 3927
```

RESULT 10
PKSM_BACSU
ID PKSM_BACSU STANDARD; PRT; 4273 AA.
AC P40872; O31781;
DT 01-FEB-1995 (Rel. 31, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Putative polyketide synthase pksM.
GN PKSM OR PKSY.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=98044033; PubMed=9384377;
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
RA Borriss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,
RA Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
RA Guiseppi G., Guy B.J., Haga K., Haiech J., Harwood C.R., Henaut A.,
RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
RA Kobayashi Y., Koetter P., Koningstein G., Krogh S., Kumano M.,
RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,
RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadaie Y.,
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,
RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA Viari A., Wambutt R., Wedler E., Wedler H., Weitzenegger T.,
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.;
RT "The complete genome sequence of the Gram-positive bacterium Bacillus
RT subtilis.";
RL Nature 390:249-256(1997).
RN [2]
RP SEQUENCE OF 1-1763 FROM N.A.
RC STRAIN=168 / PB1424;
RA Tognoni A., Grandi G.;
RL Submitted (JUL-1994) to the EMBL/GenBank/DDBJ databases.
CC -!- FUNCTION: POTENTIALLY INVOLVED IN SOME INTERMEDIATE STEPS FOR
CC THE SYNTHESIS OF A POLYKETIDE MOLECULE WHICH MAY BE INVOLVED IN
CC SECONDARY METABOLISM.
CC -!- COFACTOR: CONTAINS 4 COVALENTLY BOUND PHOSPHOPANTETHEINES
CC (POTENTIAL).
CC -!- SIMILARITY: CONTAINS 4 ACYL CARRIER DOMAINS.
CC ---
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC ---
DR EMBL; Z99113; CAB13603.1; -.
DR EMBL; Z35133; CAA84505.1; -.
DR SubtiList; BG10931; pksM.
DR InterPro; IPR002198; ADH_short.
DR InterPro; IPR000794; Ketoacyl-synt.
DR InterPro; IPR001601; Methyltransf.
DR InterPro; IPR003880; Ppantne_attach.
DR InterPro; IPR000051; SAM_bind.
DR Pfam; PF00106; adh_short; 1.
DR Pfam; PF00109; ketoacyl-synt; 3.
DR Pfam; PF00550; pp-binding; 4.
DR Pfam; PF02801; ketoacyl-synt_C; 3.
DR PROSITE; PS00012; PHOSPHOPANTETHEINE; 2.
DR PROSITE; PS00606; B_KETOACYL_SYNTHASE; 2.
DR PROSITE; PS50075; ACP_DOMAIN; 4.
KW Transferase; Acyltransferase; Antibiotic biosynthesis; NADP;
KW Phosphopantetheine; Multifunctional enzyme; Repeat; Complete proteome.
FT DOMAIN 295 364 ACYL CARRIER (ACP) 1.
FT DOMAIN 396 834 BETA-KETOACYL SYNTHASE 1.
FT DOMAIN 2190 2258 ACYL CARRIER (ACP) 2.
FT DOMAIN 2322 2737 BETA-KETOACYL SYNTHASE 2.
FT DOMAIN 3532 3947 BETA-KETOACYL SYNTHASE 3.
FT DOMAIN 3410 3483 ACYL CARRIER (ACP) 3.
FT DOMAIN 4140 4209 ACYL CARRIER (ACP) 4.
FT BINDING 327 327 PHOSPHOPANTETHEINE (POTENTIAL).
FT BINDING 2222 2222 PHOSPHOPANTETHEINE (POTENTIAL).
FT ACT_SITE 2476 2476 BETA-KETOACYL SYNTHASE (BY SIMILARITY).
FT BINDING 3446 3446 PHOSPHOPANTETHEINE (POTENTIAL).
FT ACT_SITE 3690 3690 BETA-KETOACYL SYNTHASE (BY SIMILARITY).
FT BINDING 4172 4172 PHOSPHOPANTETHEINE (POTENTIAL).
FT CONFLICT 103 103 E -> V (IN REF. 2).
FT CONFLICT 276 276 Q -> E (IN REF. 2).
FT CONFLICT 289 289 T -> S (IN REF. 2).
SQ SEQUENCE 4273 AA; 477459 MW; 3BBFCF1A250AEB5A CRC64;

 Query Match 7.9%; Score 1652; DB 1; Length 4273;
 Best Local Similarity 23.7%; Pred. No. 3.1e-67;
 Matches 614; Conservative 340; Mismatches 926; Indels 712; Gaps 76;



```
Db       3614 QPIFLEEAWHTFEDAGYMGDRIKGKSCGVYVGVEEGEYA---HLTG---DTDYINGTQNA 3667

Qy       2236 VISGRISYTFGFEGPAVSVDTACSSSLVALHLACQALRAGECSMALAGGVTVMSSPGAFV 2295

Db       3668 TLSARIAYALDLKGPNMALTAACSSGLVAIHQACSALRQGDCEMALAAGVTLNISHMSFE 3727

Qy       2296 EFSRQRGLAADGHCKAFSAAADGTGWGEGVGMLLVERLSDAHRNGHRVLAVVRGSAVNQD 2355

Db       3728 ALTRAEMLSPNGQCKVFDQDANGLVPGEAVAAVLLKPLSKAIEDKDHIYGCIKASGVNYD 3787

Qy       2356 GASNGLTAPNGPSQQRVIRQALANAGLSAGDVDAVEAHGTGTTLGDPIEAQALLATYGQD 2415

Db       3788 GKTNGITAPNPFSQAELIENIYEKNEINPLDIQYVMAHSTGSNLGDPLEVQALTSVFSKY 3847

Qy       2416 RAGEGPLWLGSVKSNVGHTQAAAGVAGVIKMVMALRHGLLPRTLHVDEPSPHVDWSAGAV 2475

Db       3848 TKQKQFCMISSIKPLIGHTFAASGTVALISMLMAMKNQIIPATHHCESENPYIPFKESPF 3907

Qy       2476 QLLTETVPWPGGEGRLRRAGVSSFGVSGTNAHVILEEAPADDVPGGPPAGEGDAGSDDEA 2535

Db       3908 VLCKENRSWIKKNQKPRMGTISTTGISGTNAHAVIEEYIPDDQP-----------STQR 3955

Qy       2536 AAGSPGVWPWLVSAKSQPALRAQAQALHAHLTDHPGLDLADVGYTLAHARAVFDHRATLI 2595

Db       3956 HQGSPQI--FVISAQNDDRLQDAACRMIAYLEQNHNLSLPDVAYTLQVGRKAMEARLAIV 4013

Qy       2596 AADRDTFLQALQ 2607

Db       4014 ANNQEQLVRKLK 4025


RESULT 11
MSAS_PENPA
ID   MSAS_PENPA     STANDARD;      PRT;  1774 AA.
AC   P22367;
DT   01-AUG-1991 (Rel. 19, Created)
DT   01-AUG-1991 (Rel. 19, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   6-methylsalicylic acid synthase (EC 2.3.1.-) (MSAS).
OS   Penicillium patulum (Penicillium griseofulvum).
OC   Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC   Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Penicillium.
OX   NCBI_TaxID=5078;
RN   [1]
RP   SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC   STRAIN=DSM 62862;
RX   MEDLINE=91006137; PubMed=2209605;
RA   Beck J., Ripka S., Siegner A., Schiltz E., Schweizer E.;
RT   "The multifunctional 6-methylsalicylic acid synthase gene of
RT   Penicillium patulum. Its gene structure relative to that of other
RT   polyketide synthases.";
RL   Eur. J. Biochem. 192:487-498(1990).
CC   -!- FUNCTION: This multifunctional enzyme is a polyketide synthase.
CC       It catalyzes a total of 11 steps by seven different component
CC       enzymes, in the biosynthesis of the antibiotic patulin.
CC   -!- CATALYTIC ACTIVITY: Acetyl-CoA + 3 malonyl-CoA + NADPH =
CC       6-methylsalicylic acid + NADP(+) + 3 CoA + 3 CO(2) + H(2)O.
CC   -!- PATHWAY: Patulin biosynthesis.
CC   -!- SUBUNIT: HOMOMULTIMER.
CC   -!- INDUCTION: IN THE LATE LOGARITHMIC GROWTH PHASE.
CC   -!- SIMILARITY: WITH MOST POLYKETIDE SYNTHASES, SPECIALLY WITH RAT
CC       FATTY ACID SYNTHASE, AND WITH OTHER ENZYMES SUCH AS LIPASES AND
CC       THIOLASES.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X55776; CAA39295.1; -.
DR   PIR; S13178; S13178.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR003880; Ppantne_attach.
DR   Pfam; PF00109; ketoacyl-synt; 1.
DR   Pfam; PF00550; pp-binding; 1.
DR   Pfam; PF00698; Acyl_transf; 1.
DR   Pfam; PF02801; ketoacyl-synt_C; 1.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; 1.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 1.
DR   PROSITE; PS50075; ACP_DOMAIN; 1.
KW   Multifunctional enzyme; Oxidoreductase; Antibiotic biosynthesis;
KW   Transferase; NADP; Phosphopantetheine.
FT   DOMAIN      186    238       ACYLTRANSFERASE (AT).
FT   DOMAIN      642    676       ACETYL/MALONYL TRANSFERASES.
FT   DOMAIN     1403   1450       2-OXOACYL REDUCTASE.
FT   DOMAIN     1700   1769       ACYL CARRIER (ACP).
FT   NP_BIND    1419   1424       NADP (POTENTIAL).
FT   ACT_SITE    204    204       BETA-KETOACYL SYNTHASE (BY SIMILARITY).
FT   ACT_SITE    653    653       MALONYLTRANSFERASE (BY SIMILARITY).
FT   BINDING    1732   1732       PHOSPHOPANTETHEINE (BY SIMILARITY).
SQ   SEQUENCE   1774 AA;  190732 MW;  05ED5DD10863F938 CRC64;

  Query Match             7.9%;  Score 1638.5;  DB 1;  Length 1774;
  Best Local Similarity  29.4%;  Pred. No. 4.3e-67;
  Matches  551;  Conservative  236;  Mismatches  699;  Indels  389;  Gaps   64;

Qy        463 TAPVPDSPSGHGGTAAAD----DPIAIIGMACRFPGGVRSPKDLWELAASGGDAIGPFPT 518

Db         14 TSP---APVGTPGTEYSEYEFSNDVAVVGMACRVAGGNHNPELLWQSLLSQKSAMGEIPP 70

Qy        519 DRGWPTEQRHA------QDPTQPGTFYPQGGGFLHDAAHFDAGFFGISPREALAMDPQQR 572

Db         71 MRWEPYYRRDARNEKFLKNTTSRGYFLDR----LED---FDCQFFGISPKEAEQMDPQQR 123

Qy        573 LLLETSWEAFERAGIDPLSVRGSRTGVFAGALSFDYGPRMDTASSEGAADVEGHILTGTT 632

Db        124 VSLEVASEALEDAGIPAKSLSGSDTAVFWGVNSDDYSKLV----LEDLPNVEAWMGIGTA 179

Qy        633 GSVLSGRIAYSFGLEGPAITVDTGCSASLVTLHLACQSLRSGECTLALAGGVSVMSTLGM 692

Db        180 YCGVPNRISYHLNLMGPSTAVDAACASSLVAIHHGVQAIRLGESKVAIVGGVNALCGPGL 239

Qy        693 FIEFSRQRGLSVDGRCKAYSAAADGTGWGEGVGMLLVERLSDAVRLGHRVLAVVRGSAVN 752

Db        240 TRVLDKAGAISSDGSCKSFDDDAHGYARGEGAGALVLKSLHRALLDHDNVLAVIKGSAVC 299

Qy        753 QDGASNGLTAPNGPAQERVIRQALANAGLSVADVDVVEGHGTGTTLGDPIEAQALLATYG 812

Db        300 QDGKTNGIMAPNSVAQQLAANNALSAANIDPHTVRYVEAHATSTPLGDPTEISAIASVYG 359

Qy        813 -QRAGDRPLWLGSLKSNIGHTMAAAGVGGVIKMVMALREGVLPRTLHVDKPSPQVDWSAG 871

Db        360 ADRPADDPCYIGSIKPNIGHLEAGAGVMGFIKAVLAIQKGVLPPQANLTKLNSRIDWKTA 419

Qy        872 AVRLLTEAVPWPGDAAGRLRRAGVSSFGIGGTNAHVILEE-APAAGGCVAGGGVLEGAPG 930

Db        420 GVKVVQEATPWP--ESDPIRRAGVCSYGYGGTVSHAVIEEFSPILQPDPLGNGAVSG-PG 476

Qy        931 LAISVAESVAAPVAVSAPVAESVPVPVPVPVPVPVSARSEAGLRAQAEALRQYV-AVRPD 989

Db        477 LLL----------------------------LSGPQEKRLALQAKTLRDWMTAEGKD 505

Qy        990 VSLADVGAGLACGRAVLEHRAVVLAADREELVQGLGALAAGEPDRRVTTGHAPGGD-RGG 1048

Db        506 HNLSDILTTLATRRDHHDYRAALVVDDYRDAEQVLQSLANGVDHTFTTQSRVLGSDISKD 565

Qy       1049 VVFVFPGQGGQWAGMGVRLLASSPVFARRMQACEEALAPWVDWSVVDILRRDAGDAVWER 1108

Db        566 VVWVFSGHGAQWPDMGKQLI-HNPVFFAAIQPLDELIQAEIGLSPIELLR--TGD--FES 620

Qy       1109 ADVVQPVLFSVMVSLAALWRSYGIEPDAVLGHSQGEIAAAHVCGALSLKDAAKTVALRSR 1168
```

```
Db        621 SDRVQILTYVMQIGLSALLQSNGITPQAVIGHSVGEIAASVVAGALSPAEGALIVTRRAL 680

Qy       1169 ALAAVRGRGGMASVPLPAQEVEQLIGERWAGRLWVAAVNGPRSTAVSGDAEAVDEVLAYC 1228
              | |:|||  | ||: | |:::| |    | ||  : | |  |:|| | | |
Db        681 LYRQVMGKGGMILVNLPSAETEEILGSR--SDLVVAIDSSPSSCVVAGDKELVAETAEAL 738

Qy       1229 AGTGVRARRIPVDYASHCPH----VQPLREELLELLGDISPQPSGVPFFSTVEGTWLDTT 1284
                ||:   :  | | | |     | |||: | | |  :||    |  :||
Db        739 KARGVKTFTVKSDIAFHSPTLNGLVDPLRDVLAETLSPVSP---NVKLYSTALADPRGQD 795

Qy       1285 TLDAAYWYRNLHQPVRFSDAVQALADDGHRVFVEVSPHPTLVPAIEDTTEDTA-EDVTAI 1343
                |  ||  |:   || : ||:|  :||:|:|:||| || :  :| :|  |   ||   |
Db        796 LRDVEYWAGNMVNRVRLTSAVKAAVEDGYRLFLEVSTHPVVSHSINETLMDAGMEDFAVI 855

Qy       1344 GSLRRGDNDTRRFLTALAHTHTTG--------------IGTP-TTWHH------------ 1376
                :| |     :  | ::|  |  |               | | ||| |
Db        856 PTLLRKKPTEKHILHSIAQLHCRGAEVNWAAQMPGRWATGVPTTTWMHKPIWRKIETAPL 915

Qy       1377 -----HYTHHHT-----HPHP-------HTHLDLPTYPFQHQHYWLESSQ--PGAG---- 1413
                   |    ||       | |         | ||  | ||   |  |  ::  | ||
Db        916 HTGLTHDVEKHTLLGQRIPVPGTDTYVYTTRLDNDTKPFPGSHP-LHGTEIVPAAGLINT 974

Qy       1414 --SGSGAG-----------AGSGAGSGRAGTAGGTAEVESR------------------- 1441
                 |:|              | :   | :          :| ||
Db        975 FLKGTGGQMLQNVVLRVPVAINAPRSVQVVVQQDQVKVVSRLIPSEPSQLDDDASWVTHT 1034

Qy       1442 --FWD-----AVARQDLETVATTLAVPPSAG-----LDTV-VPALS-AW---HRHQHDQ- 1483
                 :||     :  | |   | : |   :        || | | |:    |      :::|:
Db       1035 TAYWDRKVAGSEDRIDFAAVKSRLVTKLADNFSIDYLDKVGVSAMGFPWAVTEHYRNDKE 1094

Qy       1484 --ARINT-----------WTYQETWKPL------------TLPTTHQPHQ---------- 1508
                 ||::             |   :| |:             |    | |
Db       1095 MLARVDVNPAISGDAPLPWD-SSSWAPVLDAATSVGSTIFPTPALRMPAQIERVEVFTSQ 1153

Qy       1509 -----TWLI---AIPETQTHH-------------------------PHITNILTNLHHH- 1534
                    :||    |     | |                        | ::  : :| |
Db       1154 DPPKISWLYVQEASDSVPTSHVSVVSEAGEVLAKFTAMRFSEIEGTPGVSGSMESLVHQI 1213

Qy       1535 ---GITPI--PLTLN--------------HTHTNPQHLHHTLHHTRQQ------------ 1563
                   ||   ||::.                :  : |  ::     : |:
Db       1214 AWPPATPAEEPLSIETVILVSPDATTRALYAASLPTRVNSFQFSSTQEFFSNASSLPLEK 1273

Qy       1564 --AQNHTTGAITGLLSLLALDETPHPHHPHTPTGTLLNL--------------TLTQTHT 1607
                   | |  |  | |  |         : |  || |                 |||
Db       1274 GTVVTYIPGEVASLAEVPAASE--------SFTWNLLELIKFTVNGSLPIKVFTLTANIG 1325

Qy       1608 QTHPPTPLWYATTNATTTHPNDPLTHPTQAQTWGLARTTLLEHPTHTAGIIDLPTTPTPH 1667
                :   ||.|                  |:  :|||| |||     :||:     |
Db       1326 EGQTPTAL------------------AQSPLYGLARVIASEHP-DLGTLIDVEEPVIPL 1365

Qy       1668 TLQHLTQTLTQPHHQTQLAIRTTGTHTRRLTPTTLTPTHQPP--TPTPHGTTLITGGTGA 1725
                :  |          :| ||  :  | |    | |  : |   | | || |||||
Db       1366 STMRYIQGADIIRINDGIA-RT--SRFRSLPRNKLLPASEGPRLLPRPEGTYLITGGLGV 1422

Qy       1726 LATHLTHHLTTHQPTQHLLLTSRTGPHTPHAQHLTTQLQ----------QKGIHLTITTC 1775
                |   :   |      : ||:: |  |        ::  ||             :|  : :
Db       1423 LGLEVADFLVEKGARRLLLISRRALPPRRTWDQVSEDLQPTIAKIRLLESRGASVHVLPL 1482

Qy       1776 DTSNPDQLQQLLN-----TIPPQHPLTTVIHTAGILDDATLTNLTPTQLNNVLRAKAHSA 1830
                | : || ::||       ::|    :  |:| ||:||:  :   |   | ||  |   |
Db       1483 DITKPDAVEQLTTALDRLSLP---SVQGVVHAAGVLDNELVMQTTRDAFNRVLAPKIAGA 1539

Qy       1831 HLLHQLTQHTPLTAFVLYSSAAATFGAPGQANYAAANAYLDALAHHRHTHHLPATSIAWG 1890
                  ||::     :  ||:::||     |  |||:| : ||:|| || ||      | |  |
Db       1540 LALHEVFPPKSVDFFVMFSSCGNLVGFTGQASYGSGNAFLDTLATHRARLGDAAVSFQWT 1599

Qy       1891 TWQGNGL-ADSDKARAYLDRRGFRPMSPELATAA---------------VTQAIADTER- 1933
                :|:|  |: | :|    | |: :|   :: : | ||               ::|  | |
Db       1600 SWRGLGMGASTDFINAELESKGITDVTRDEAFAAWQHLAKYDMDHGVVLRSRAFEDGEPI 1659

Qy       1934 PYVVIADIDWSKIEHTSQTSDLVSAARAEREPAVQRPTPPAELHKTLAHQTSADQRAALLE 1993
                |  :: ||   ::   | ||    ||     ||  ||   ||          :| | |
Db       1660 PVSILNDIAVRRVGTVSNTS---PAAAGSSDAV--PTSGPEL------------KAYLDE 1702

Qy       1994 LVRDHVAAVLR----HADPKAIAPDQSFRALGFDSLTAVEFRNLLIKATGLRLPVSLVFD 2049
                  :|  || ||:      | ||   |      || ||:  |  | |     : :|.:| :
Db       1703 KIRGCVAKVLQMTAEDVDSKAALAD-----LGVDSVMTVTLRRQLQLTLKIAVPPTLTWS 1757

Qy       2050 HPTPAKLAVHLQNQL 2064
                 ||| : |||    :|
Db       1758 HPTVSHLAVWFAEKL 1772
```

```
RESULT 12
PKS1_ASPPA
ID   PKS1_ASPPA     STANDARD;      PRT;  2109 AA.
AC   Q12053;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Aflatoxin biosynthesis polyketide synthase (PKS).
GN   PKSL1.
OS   Aspergillus parasiticus.
OC   Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC   Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
OX   NCBI_TaxID=5067;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NRRL 2999;
RX   MEDLINE=96042102; PubMed=7592391;
RA   Feng G.H., Leonard T.J.;
RT   "Characterization of the polyketide synthase gene (pksL1) required
RT   for aflatoxin biosynthesis in Aspergillus parasiticus.";
RL   J. Bacteriol. 177:6246-6254(1995).
CC   -!- FUNCTION: INVOLVED IN THE SYNTHESIS OF THE POLYKETIDE NUCLEUS OF
CC       AFLATOXIN FROM HEXANOYL COA AND SEVEN MALONATES.
CC   -!- COFACTOR: CONTAINS 1 COVALENTLY BOUND PHOSPHOPANTETHEINE (By
CC       similarity).
CC   -!- PATHWAY: Aflatoxin biosynthesis; first step.
CC   -!- SIMILARITY: CONTAINS 1 ACYL CARRIER DOMAIN.
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; L42766; AAC41675.1; -.
DR   EMBL; L42765; AAC41674.1; -.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR003880; Ppantne_attach.
DR   InterPro; IPR001031; Thioesterase.
DR   Pfam; PF00109; ketoacyl-synt; 1.
DR   Pfam; PF00550; pp-binding; 1.
DR   Pfam; PF00698; Acyl_transf; 1.
DR   Pfam; PF00975; Thioesterase; 1.
DR   Pfam; PF02801; ketoacyl-synt_C; 1.
DR   PROSITE; PS50075; ACP_DOMAIN; 1.
KW   Transferase; Acyltransferase; Phosphopantetheine;
KW   Multifunctional enzyme.
FT   DOMAIN      374    805       BETA-KETOACYL SYNTHASE.
FT   DOMAIN        ?      ?       ACYL/MALONYL TRANSFERASES.
FT   DOMAIN     1714   1785       ACYL CARRIER (ACP).
FT   DOMAIN        ?   2109       THIOESTERASE.
FT   ACT_SITE    543    543       BETA-KETOACYL SYNTHASE (BY SIMILARITY).
FT   ACT_SITE    993    993       ACYL/MALONYL TRANSFERASES (BY
FT                                SIMILARITY).
FT   BINDING    1746   1746       PHOSPHOPANTETHEINE (BY SIMILARITY).
FT   ACT_SITE   1937   1937       THIOESTERASE (BY SIMILARITY).
SQ   SEQUENCE   2109 AA;  230715 MW;  CB701372A16D8551 CRC64;

  Query Match           5.4%;  Score 1117.5;  DB 1;  Length 2109;
  Best Local Similarity  24.6%;  Pred. No. 2.8e-43;
```

```
  Matches  448;  Conservative  229;  Mismatches  718;  Indels  423;  Gaps   55;

Qy       1919 LATAAVTQAIADT----------ERPYVVIADIDWSKIEHTSQTSDLVSAARAREPAVQ 1967
Db        249 LFTQADITTILDTTPTTPWEQLPGQIPYISHVT---GNVVQTSNYRDLIEVA-------- 297

Qy       1968 RPTPPAELHKTLAHQTSADQRAALLELVRDHVAAVLRHADPKAIAPDQSFRALGFDSLTA 2027
Db        298 -------LSETLLEQVR-------LDLVETGLPRLLQSRQVK-------------SVTI 329

Qy       2028 VEFRNLLIKATGLRLPVSLVFDHPTPAKLAVHLQNQLRGTAAESAPSAAAVTAEASVTEP 2087
Db        330 VPFLTRMNETMSNILPDSPI------------STETRTDTGRAIPASGRPGAGKC---K 373

Qy       2088 IAIVGMACRFPGGVTSADDFWDLISSEQDAIGGFPTDRGWDLDTLYDPDPDHPGTCYTRN 2147
Db        374 LAIVSMSGRFPESPTT-ESFWDLLYKGLDVCKEVPR-RRWDINTHVDPSGKARNKGATKW 431

Qy       2148 GGFLYDAGHFDAEFFGISPREALAMDPQQRLLLETAWETIEHAGINPHTLHGTP---TGV 2204
Db        432 GCWLDFSGDFDPRFFGISPKEAPQMDPAQRMALMSTYEAMERAGLVPDTTPSTQRDRIGV 491

Qy       2205 FTGTNGQDYALRVHNAGQSTDGFALTGTAGSVISGRISYTFGFEGPAVSVDTACSSSLVA 2264
Db        492 FHGVTSNDW--METNTAQNIDTYFITGGNRGFIPGRINFCFEFAGPSYTNDTACSSSLAA 549

Qy       2265 LHLACQALRAGECSMALAGGVTVMSSPGAFVEFSRQRGLAADGHCKAFSAAADGTGWGEG 2324
Db        550 IHLACNSLWRGDCDTAVAGGTNMIYTPDGHTGLDKGFFLSRTGNCKPYDDKADGYCRAEG 609

Qy       2325 VGMLLVERLSDAHRNGHRVLAVVRGSAVNQDGASNGLTAPNGPSQQRVIRQALANAGLSA 2384
Db        610 VGTVFIKRLEDALADNDPILGVILDAKTNHSAMSESMTRPHVGAQIDNMTAALNTTGLHP 669

Qy       2385 GDVDAVEAHGTGTTLGDPIEAQALLATYGQD---RAGEGPLWLGSVKSNVGHTQAAAGVA 2441
Db        670 NDFSYIEMHGTGTQVGDAVEMESVLSVFAPSETARKADQPLFVGSAKANVGHGEGVSGVT 729

Qy       2442 GVIKMVMALRHGLLPRTLHVDEPSPHV------------DWSAGAVQLLTETVPWPGGE 2488
Db        730 SLIKVLMMMQHDTIP---------PHCGIKPGSKINRNFPDLGARNVHIAFEPKPWPRTH 780

Qy       2489 GRLRRAGVSSFGVSGTNAHVILEEAPADDVPGGPPAGEGDAGSDDEAAAGSPGVWPWLVS 2548
Db        781 -TPRRVLINNFSAAGGNTALIVEDAPERHWP-----TEKDPRSSHIVA----------LS 824

Qy       2549 AKSQPALRAQAQALHAHLTDHPGLDLADVGYTLAHARAVFDHRATLIAADRDTFLQALQ- 2607
Db        825 AHVGASMKTNLERLHQYLLKNPHTDLAQLSYTTTARRWHYLHRVSVTGASVEEVTRKLEM 884

Qy       2608 ALAAGEPHPAVIHSSAPGGTGTGEAAGKTAFICSGQGTQRPGMAHGLYHTHPVFAAALND 2667
Db        885 AIQNGD------------GVSRPKSKPKILFAFTGQGSQYATMGKQVYDAYPSFREDLEK 932

Qy       2668 ICTHLDPHLDHPLLPLLTQNDNDNEDAAALLQQTRYAQPALFAFQVALHRLLTDGYHITP 2727
Db        933 FDRLAQSHGFPSFLHVCTSPKGDVEEMAPVV-----VQLAITCLQMALTNLMT-SFGIRP 986

Qy       2728 HYYAGHSLGEITAAHLAGILTLTDATTLITQRATLMQ-TMPPGT--MTTLHTTPHHITHH 2784
Db        987 DVTVGHSLGEFAALYAAGVLSASDVVYLVGQRAELLQERCQRGTHAMLAVKATPEALSQW 1046

Qy       2785 LTAHENDLAIAAINTPTSLVISGTPHTVQHITTLCQQQGIKTKTLPTNHAFHSPHTNPIL 2844
Db       1047 IQDH--DCEVACINGPEDTVLSGTTKNVAEVQRAMTDNGIKCTLLKLPFAFHSAQVQPIL 1104

Qy       2845 NQLHQHTQTLTYHPPH----TPLITANTPPDQLLTPHYWTQQARNTVDYATTTQTLHQHG 2900
Db       1105 DDFEALAQGATFAKPQLLILSPLLRTEIHEQGVVTPSYVAQHCRHTVDMAQALRSAREKG 1164

Qy       2901 V----TTYIELGPDNTLTTLTHHNLPNPPTTTLTLTHPHHHPQTHLLTNLAKTTTT---- 2952
Db       1165 LIDDKTLVIELGPKPLISGMVKMTLGDKISTLPTLA-----PNKAIWPSLQKILTSVYTG 1219

Qy       2953 -WHPHHYTHHDNQPHTHTHLDLPTY---------PFQ------HHHYWLESTQPGAGNVS 2996
Db       1220 GWDINWKKYHAPFASSQKVVDLPSYGWDLKDYYIPYQGDWCLHRHQQDCKCAAPGHEIKT 1279

Qy       2997 A-------------AGLDPTEHPL-------------------LGATLELATDGGALLAG 3024
Db       1280 ADYQVPPESTPHRPSKLDPSKEAFPEIKTTTTLHRVVEETTKPLGATLVVETD------- 1332

Qy       3025 RLSLRSHPWLA-DHAVGGTVLLSGATFLELALHAGTYVGCDRVDELTLHAPLVVPVDGGV 3083
Db       1333 -ISRKDVNGLARGHLVDGIPLCTPSFYADIAMQVGQY----SMQRLRAGHPGAGAIDGLV 1387

Qy       3084 SVQVGVA----AADGEGRRLVSVYARGGSACGGGGASGGVWTCHASGVLVEAAAGGVVVD 3139
Db       1388 DVSDMVVDKALVPHGKGPQLLRT---------------------------------- 1410

Qy       3140 GLAGVWPPRGAVAVD----------VDGVRDRLAGAGCVLGPVFSGLRAVWRDGGD---- 3185
Db       1411 TLTMEWPPKAAATTRSAKVKFATYFADGKLDTEHASCTVRFTSDAQLKSLRRSVSEYKTH 1470

Qy       3186 ----------------------------------LLAEVCLPEEAWGDAAG-------- 3202
Db       1471 IRQLHDGHAKGQFMRYNRKTGYKLMSSMARFNPDYMLLDYLVLNEAENEAASGVDFSLGS 1530

Qy       3203 ----FGLHPALLDGVVQPLSVLLPGGTGFGEGAG----FGEGVRVPAVWGGVSLHRAGVT 3254
Db       1531 SEGTFAAHPAHVDAITQV--------AGFAMNANDNVDIEKQVYVNHGWDSFQIYQPLDN 1582

Qy       3255 GVRVRV-SAVGRGGGREAV--SVVVGDEAGVPVASVDRLELRPVDMGQLRAVSVSAGRRG 3311
Db       1583 SKSYQVYTKMGQAKENDLVHGDVVVLDGEQI-VAFFRGLTLRSVPRGALRVVLQTTVKKA 1641

Qy       3312 ---------------------SLYAVQWAEVGPVPVCGQAWAWHEDVGESGGGPVPGVVV 3350
Db       1642 DRQLGFKTMPSPPPPTTTMPISPYKPANTQVSSQAIPAEATHSHTP-PQPKHSPVPET-- 1698

Qy       3351 LRCPDAGAGGGGGGGGGGVGEVVGGVLGVVQGWLGL--------ERFAG------SRLV 3396
Db       1699 -------AGSAPAAKGVGVSNEKLDAVMRVVSEESGIALEELTDDSNFADMGIDSLSSMV 1751

Qy       3397 VVTRGAVVAGPEDGP-----VD----------VVGASVWGLVRSAQAEHPDRFVLLDLDT 3441
Db       1752 IGSRFREDLGLDLGPEFSLFIDCTTVRALKDFMLGSGDAGSGSNVEDPPPSATPGINPET 1811

Qy       3442 DTGTDL--------DTGAGAGWGVDGGRVAA-----------VVACGEPQLAVR----- 3476
Db       1812 DWSSSASDSIFASEDHGHSSESGADTGSPPALDLKPYCRPSTSVVLQGLPMVARKTLFML 1871

Qy       3477 ----GERLLAARLKRLES 3490
Db       1872 PDGGGSAFSYASLPRLKS 1889
```

RESULT 13
WA_EMENI
ID WA_EMENI STANDARD; PRT; 1986 AA.
AC Q03149;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Conidial green pigment synthase (EC 2.3.1.-).
GN WA.
OS Emericella nidulans (Aspergillus nidulans).
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eurotiales; Trichocomaceae; Emericella.
OX NCBI_TaxID=5072;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93101122; PubMed=1465094;
RA Mayorga M.E., Timberlake W.E.;
RT "The developmentally regulated Aspergillus nidulans wA gene encodes a
RT polypeptide homologous to polyketide and fatty acid synthases.";

```
RL   Mol. Gen. Genet. 235:205-212(1992).
CC   -!- FUNCTION: THIS PROTEIN CONDENSES CARBON UNITS TO FORM AN
CC       INTERMEDIATE YELLOW POLYKETIDE PIGMENT THAT IS POLYMERIZED
CC       BY CONIDIAL LACCASE TO FORM THE GREEN PIGMENT IN MATURE
CC       ASEXUAL SPORES (CONIDIA).
CC   -!- COFACTOR: CONTAINS 2 COVALENTLY BOUND PHOSPHOPANTETHEINES
CC       (POTENTIAL).
CC   -!- PATHWAY: Conidial green pigment biosynthesis.
CC   -!- SIMILARITY: CONTAINS 2 ACYL CARRIER DOMAINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X65866; CAA46695.1; -.
DR   PIR; S28353; S28353.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR003880; Ppantne_attach.
DR   InterPro; IPR001031; Thioesterase.
DR   Pfam; PF00109; ketoacyl-synt; 1.
DR   Pfam; PF00550; pp-binding; 2.
DR   Pfam; PF00698; Acyl_transf; 1.
DR   Pfam; PF00975; Thioesterase; 1.
DR   Pfam; PF02801; ketoacyl-synt_C; 1.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; 1.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 1.
DR   PROSITE; PS50075; ACP_DOMAIN; 2.
KW   Transferase; Phosphopantetheine; Multifunctional enzyme; Repeat.
FT   DOMAIN      529    582       BETA-KETOACYL SYNTHASE (BY SIMILARITY).
FT   DOMAIN      991   1024       ACYL/MALONYL TRANSFERASES
FT                                (BY SIMILARITY).
FT   DOMAIN     1650   1719       ACYL CARRIER (ACP) 1.
FT   DOMAIN     1772   1841       ACYL CARRIER (ACP) 2.
FT   ACT_SITE    548    548       BETA-KETOACYL SYNTHASE (BY SIMILARITY).
FT   ACT_SITE   1001   1001       ACYL/MALONYL TRANSFERASES (BY
FT                                SIMILARITY).
FT   BINDING    1682   1682       PHOSPHOPANTETHEINE (BY SIMILARITY).
FT   BINDING    1804   1804       PHOSPHOPANTETHEINE (BY SIMILARITY).
SQ   SEQUENCE   1986 AA;  216634 MW;  74EF0940FF40EE9A CRC64;

  Query Match             5.1%;  Score 1064.5;  DB 1;  Length 1986;
  Best Local Similarity   27.9%;  Pred. No. 6.7e-41;
  Matches  324;  Conservative  189;  Mismatches  488;  Indels  160;  Gaps   37;

Qy       2034 LIKATGLRLPVSLVFDHPTPAKLAVHLQNQLR----------GTAAESAPSAAAV--TAE 2081
                 ::|: |   |  |:   |  :     | | |:          |  :: |     |  |
Db        317 ILKSVGDNKPKKLL---PISSTATQSLFNSLKKSNLVNIEVDGGISDFAAETQLVNQTGR 373

Qy       2082 ASVTEPIAIVGMACRFPGGVTSADDFWDLISSEQDAIGGFPTDRGWDLDTLYDPDPDHPG 2141
                | ::: |||:||: |||    |  |||:|:    |     | || || |   |
Db        374 AELSK-IAIIGMSGRFP-EADSPQDFWNLLYKGLDVHRKVPEDR-WDADAHVDLTGTATN 430

Qy       2142 TCYTRNGGFLYDAGHFDAEFFGISPREALAMDPQQRLLLETAWETIEHAGINPHTLHGTP 2201
                |      | :: : | ||  || :||||||  || ||| | ||:| :| ||  | :   |
Db        431 TSKVPYGCWIREPGLFDPRFFNMSPREALQADPAQRLALLTAYEALEGAGFVPDSTPSTQ 490

Qy       2202 ---TGVFTGTNGQDYALRVHNAGQSTDGFALTGTAGSVISGRISYTFGFEGPAVSVDTAC 2258
                   |:| |    ||  |  |:||  | : : |   :    |||:| | | ||:|||||
Db        491 RDRVGIFYGMTSDDY--REVNSGQDIDTYFIPGGNRAFTPGRINYYFKFSGPSVSVDTAC 548

Qy       2259 SSSLVALHLACQALRAGECSMALAGGVTVMSSPGAFVEFSRQRGLAADGHCKAFSAAADG 2318
                 |||| |:|||| ::   :|  |: ||| ::::|       |   |:  |:|  |   |||
Db        549 SSSLAAIHLACNSIWRNDCDTAITGGVNILTNPDNHAGLDRGHFLSRTGNCNTFDDGADG 608

Qy       2319 TGWGEGVGMLLVERLSDAHRNGHRVLAVVRGSAVNQDGASNGLTAPNGPSQQRVIRQALA 2378
                    :||| ::::|| ||  :   :| |: |:  |    :  :| |:  :|  : :: |
Db        609 YCRADGVGTVVLKRLEDALADNDPILGVINGAYTNHSAEAVSITRPHVGAQAFIFKKLLN 668

Qy       2379 NAGLSAGDVDAVEAHGTGTTLGDPIEAQALLATYGQD-RAGEG-PLWLGSVKSNVGHTQA 2436
                |  :   ::  :| |||||  || :| |::|  :  | | | |  | ||| |||:|| ::
Db        669 EANVDPKNISYIEMHGTGTQAGDAVEMQSVLDVFAPDHRRGPGQSLHLGSAKSNIGHGES 728

Qy       2437 AAGVAGVIKMVMALRHGLLPRTLHVDEPSPH---VDWSAGAVQLLTETVPWPGGEGRLRR 2493
                |:||  ::|::: ::  ::|    :     |    | :   | :  :   |       |:
Db        729 ASGVTSLVKVLLMMKENMIPPHCGIKTKINHNFPTDLAQRNVHIALQPTAWNRPSFGKRQ 788

Qy       2494 AGVSSFGVSGTNAHVILEEAPADDVPGGPPAGEGDAGSDDEAAAGSPGVWPWLVSAKSQP 2553
                  :::|  :| |  ::||: |  |     | ||   :                :||:||
Db        789 IFLNNFSAAGGNTALLLEDGPVSD-----PEGEDKRRTH-----------VITLSARSQT 832

Qy       2554 ALRAQAQALHAHLTDHP---GLD----LADVGYTLAHARAVFDHRATLIAADRDTFLQAL 2606
                ||:    ||  ::::     |:     |  : ||    |      | | | :      :|
Db        833 ALQNNIDALCQYISEQEKTFGVKDSNALPSLAYTTTARRIHHPFRVTAIGSSFQEMRDSL 892

Qy       2607 QALAAGEPHPAVIHSSAPGGTGTGEAAGKTAFICSGQGTQRPGMAHGLYHTHPVFAAALN 2666
                  :|:       : :  ||           |: :||| |   |   ||     | :|:
Db        893 --IASSRKEFVAVPAKTPG----------IGFLFTGQGAQYAAMGKQLYEDCSHFRSAIE 940

Qy       2667 DICTHLD---PHLDHP-LLPLLTQNDNDNEDAAALLQ-QTRYAQPALFAFQVALHRLLTD 2721
                   |||      | | :|||:  :   :| : ::|  |   | || :|  :|
Db        941 ----HLDCISQGQDLPSILPLVDGSLPLSELSPVVVQLGTTCVQMALSSFWASL------ 990

Qy       2722 GYHITPHYYAGHSLGEITAAHLAGILTLTDATTLITQRATLM-QTMPPGTMTTLHTTPHH 2780
                  ||| :  |||||:  | : ||:|: :|     :|| |: :  |||   |
Db        991 --GITPSFVLGHSLGDFAAMNAAGVLSTSDTIYACGRRAQLLTERCQPGTHAMLAIKAPL 1048

Qy       2781 ITHHLTAHENDLAIAAINTPTSLVISGTPHTVQHITTLCQQQGIKTKTLPTNHAFHSPHT 2840
                :      :|    :| ||:|:  ||||   ::  ::  | ::|:|:  |   :||||
Db       1049 VEVKQLLNEKVHDMACINSPSETVISGPKSSIDELSRACSEKGLKSTILTVPYAFHSAQV 1108

Qy       2841 NPILNQLHQHTQTLTYHPPHTPLITA------NTPPDQLLTPHYWTQQARNTVDYATTTQ 2894
                  |||  | :  | :|:: |  | ::|          :|   |  :  | ||:: :  :
Db       1109 EPILEDLEKALQGITFNKPSVPFVSALLGEVITEAGSNILNAEYLVRHCRETVNFLSAFE 1168

Qy       2895 TLHQHGV----TTYIELGPDNTLTTLTHHNLPNPPTTTLTLTHPHHHPQTHLLTN----- 2945
                :    :    | ::|:||    : :    |  | |||:          :|:|
Db       1169 AVRNAKLGGDQTLWLEVGPHTVCSGMVKATL-GPQTTTMASLR-RDEDTWKVLSNSLSSL 1226

Qy       2946 -LAKTTTTWHPHHYTHHDNQPHTHTHLDLPTYPFQHHHYWLE------STQPGAGNVSAA 2998
                 ||      |      :| :   :|  | |||| :   :||:          |:  : : ::|
Db       1227 YLAGVDINWK----QYHQDFSSSHRVLPLPTYKWDLKNYWIPYRNNFCLTKGSSMSAASA 1282

Qy       2999 GLDPTEHPLLGATLELATDGGALLAGR-----LSLRSHPWLAD---------HAVGGTVL 3044
                 | ||          | |    ::  |      ::  |  :||           | | |  |
Db       1283 SLQPT---------FLTTSAQRVVESRDDGLTATVVVHNDIADPDLNRVIQGHKVNGAAL 1333

Qy       3045 LSGATFLELALHAGTYV-----------GCDRVDELTLHAPLVVP--------------V 3079
                 : : :|      |:                | | |  :|:  ||:           |
Db       1334 CPSSLYADSAQTLAEYLIEKYKPELKGSGLD-VCNVTVPKPLIAKTGKEQFRISATANWV 1392

Qy       3080 DGGVSVQVGVAAADGEGRRLV 3100
                 |  |||||    |  ||::|:
Db       1393 DKHVSVQVFSVTA--EGKKLI 1411


RESULT 14
STCA_EMENI
ID   STCA_EMENI     STANDARD;      PRT;  2181 AA.
AC   Q12397;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Putative sterigmatocystin biosynthesis polyketide synthase (PKS).
GN   STCA OR PKSST.
OS   Emericella nidulans (Aspergillus nidulans).
OC   Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC   Eurotiales; Trichocomaceae; Emericella.
OX   NCBI_TaxID=5072;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=FGSC 26;
```

```
RX   MEDLINE=96202293; PubMed=8643646;
RA   Brown D.W., Yu J.-H., Kelkar H.S., Fernandes M., Nesbitt T.C.,
RA   Keller N.P., Adams T.H., Leonard T.J.;
RT   "Twenty-five coregulated transcripts define a sterigmatocystin gene
RT   cluster in Aspergillus nidulans.";
RL   Proc. Natl. Acad. Sci. U.S.A. 93:1418-1422(1996).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=FGSC 4;
RX   MEDLINE=95370159; PubMed=7642507;
RA   Yu J.-H., Leonard T.J.;
RT   "Sterigmatocystin biosynthesis in Aspergillus nidulans requires a
RT   novel type I polyketide synthase.";
RL   J. Bacteriol. 177:4792-4800(1995).
CC   -!- FUNCTION: INVOLVED IN THE SYNTHESIS OF THE POLYKETIDE NUCLEUS OF
CC       STERIGMATOCYSTIN FROM HEXANOYL-COA AND SEVEN MALONATES.
CC   -!- COFACTOR: CONTAINS 2 COVALENTLY BOUND PHOSPHOPANTETHEINES
CC       (POTENTIAL).
CC   -!- PATHWAY: Sterigmatocystin biosynthesis; first step.
CC   -!- SIMILARITY: CONTAINS 2 ACYL CARRIER DOMAINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U34740; AAC49191.1; -.
DR   EMBL; L39121; AAA81586.1; -.
DR   HSSP; P25715; 1MLA.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR003880; Ppantne_attach.
DR   InterPro; IPR001031; Thioesterase.
DR   Pfam; PF00109; ketoacyl-synt; 1.
DR   Pfam; PF00550; pp-binding; 2.
DR   Pfam; PF00698; Acyl_transf; 1.
DR   Pfam; PF00975; Thioesterase; 1.
DR   Pfam; PF02801; ketoacyl-synt_C; 1.
DR   PROSITE; PS50075; ACP_DOMAIN; 2.
KW   Transferase; Acyltransferase; Phosphopantetheine; Repeat;
KW   Multifunctional enzyme.
FT   DOMAIN      383    814       BETA-KETOACYL SYNTHASE.
FT   DOMAIN      884   1209       ACYL/MALONYL TRANSFERASES.
FT   DOMAIN     1706   1777       ACYL CARRIER (ACP) 1.
FT   DOMAIN     1830   1901       ACYL CARRIER (ACP) 2.
FT   DOMAIN        ?   2181       THIOESTERASE.
FT   ACT_SITE    552    552       BETA-KETOACYL SYNTHASE (BY SIMILARITY).
FT   ACT_SITE    978    978       ACYL/MALONYL TRANSFERASES (BY
FT                                SIMILARITY).
FT   BINDING    1738   1738       PHOSPHOPANTETHEINE (BY SIMILARITY).
FT   BINDING    1862   1862       PHOSPHOPANTETHEINE (BY SIMILARITY).
FT   ACT_SITE   2028   2028       THIOESTERASE (BY SIMILARITY).
SQ   SEQUENCE   2181 AA;  238831 MW;  5A3E5712AA9AD942 CRC64;

  Query Match             4.9%;  Score 1019.5;  DB 1;  Length 2181;
  Best Local Similarity   23.9%;  Pred. No. 8.5e-39;
  Matches  412;  Conservative  219;  Mismatches  633;  Indels  461;  Gaps   56;

Qy       1993 ELVRDHVAAVLRHADPKAIAPDQSFRALGFDSLTAVEFRNLLIKATGLRLPVSLVFDHPT 2052
                 ::|   |  :|:  |              | |:|| | |  :| |:    |  :|| :
Db        315 DVVEAEVPRLLKDRD------------GLDTLTIVAFTTVLSKS----LSNALVTEGIK 357

Qy       2053 PAKLAVHLQNQLRGTAAESAPSAAAVTAEASVTEP------IAIVGMACRFPGGVTSADD 2106
                 ||:              |::   | |      |      :||| |: |||    | |
Db        358 PAE---------------PPTSINKTPERYSHRPGSDRGKLAIVSMSGRFP-EAPSTDS 400

Qy       2107 FWDLISSEQDAIGGFPTDRGWDLDTLYDPDPDHPGTCYTRNGGFLYDAGHFDAEFFGISP 2166
                 ||||:    |      |  | ||: |  ||          || | :|  || || || |||
Db        401 PWDLLYKGLDVCKEVPL-RRWDVKTHVDPSGKARNKGATRWGCWLDFAGEFDPRFFSISP 459

Qy       2167 REALAMDPQQRLLLETAWETIEHAGINPHTLHGTP---TGVFTGTNGQDYALRVHNAGQS 2223
                 :|| ||| ||: | : :| :|  || | |    |       ||| |    |:     | |:
Db        460 KEAPQMDPAQRMALMSTYEAMERGGIVPDTTPSTQRNRIGVFHGVTSNDW--METNTAQN 517

Qy       2224 TDGFALTGTAGSVISGRISYTFGFEGPAVSVDTACSSSLVALHLACQALRAGECSMALAG 2283
                 | : :||     | |||:: | | ||: | |||||||| |:|||| :|  |:|  |:||
Db        518 IDTYFITGGNRGFIPGRINFCFEFSGPSYSNDTACSSSLAAIHLACNSLWRGDCDTAVAG 577

Qy       2284 GVTVMSSPGAFVEFSRQRGLAADGHCKAFSAAADGTGWGEGVGMLLVERLSDAHRNGHRV 2343
                 |  :: :|       :   |:  |:|||  ||||    |||| : ::|| ||      :
Db        578 GTNMIFTPDGHTGLDKGFFLSRTGNCKAFDDAADGYCRAEGVGTVFIKRLEDALAENDPI 637

Qy       2344 LAVVRGSAVNQDGASNGLTAPNGPSQQRVIRQALANAGLSAGDVDAVEAHGTGTTLGDPI 2403
                 || :      |     |: :| |  |:|    :   |: ||:|  |:  :| ||||| :|| :
Db        638 LATILDIKTNHSAMSDSMTRPFKPAQIDNMSALLSTAGISPLDLSYIEMHGTGTQVGDAV 697

Qy       2404 EAQALLATYGQD---RAGEGPLWLGSVKSNVGHTQAAAGVAGVIKMVMALRHGLLPRTLH 2460
                 | :::|: :  |   |  : ||::|| |:|:|| :  :||  :||::: :::  :|
Db        698 EMESVLSLFAPDETFRPRDKPLYVGSAKANIGHGEGVSGVTSLIKVLLMMKNDTIPPHCG 757

Qy       2461 VDEPSPHV-----DWSAGAVQLLTETVPWPGGEGRLRRAGVSSFGVSGTNAHVILEEAPA 2515
                 : :|  :     |  | | :  |  |||  :   ||  :::|  :| |  |::|:||
Db        758 I-KPGSRINRNYPDLPARNVHIAFEPKPWPRTD-TPRRVLINNFSAAGGNTAVLVEDAPV 815

Qy       2516 DDVPGGPPAGEGDAGSDDEAAAGSPGVWPWLVSAKSQPALRAQAQALHAHLTDHPGLDLA 2575
                    |  ||  |||  |||                                    |  :
Db        816 ------PELGE-IAGSFGEAA---------------------------------GNQSS 834

Qy       2576 DVGYTLAHARAVFDHRATLIAADRDTFLQALQ-ALAAGEPHPAVIHSSAPGGTGTGEAAG 2634
                 :: ||     |    || ::  |:    |: :: |:| |              |
Db        835 ELSYTTTARRWHHPHRVSITGANTMEILRNVESAIARGH-----------GVNRPATKP 882

Qy       2635 KTAFICSGQGTQRPGMAHGLYHTHPVFAAALN---------DICTHLDPHLDHP------ 2679
                 |    |||||:|  ||   ||:::| | : |              : |: :  |
Db        883 KIVIACSGQGSQYTGMGWQLYNSYPTFRSDLERFDQLARSYGFPSFLEVYTSKPVGDSME 942

Qy       2680 -LLPLLTQNDNDNEDAAALLQQTRYAQPALFAFQVALHRLLTDGYHITPHYYAGHSLGEI 2738
                  |||::                  | || : ::||  ||   : : |    ||||||
Db        943 DLLPVI------------------VQLALVSLEMALGNLL-GSFGLKPSAVIGHSLGEY 982

Qy       2739 TAAHLAGILTLTDATTLITQRATLMQ-TMPPGTMTTLHTTPHHITHHLTAHENDLAIAAI 2797
                 | :::|:|:  |   |:  || |:|      ||   |       :|      |::  :|
Db        983 AALYISGVLSAADTLYLVGMRAKLLQERCQRGTHAMLAVRASPVTLCEVLAESNCEVACH 1042

Qy       2798 NTPTSLVISGTPHTVQHITTLCQQQGIKTKTLPTNHAFHSPHTNPILNQLHQHTQTLTYH 2857
                 | |   |:||    | ::       |||   |    ||||    ||| :      : :|:|
Db       1043 NGPNDTVLSGPLKEVMNLQNSLSATGIKGTLLKLPFAFHSAQVQPILEEFKNVARGVTFH 1102

Qy       2858 PPH----TPLITANTPPDQLLTPHYWTQQARNTVDYATTTQ-TLHQHGV---TTYIELGP 2909
                    |      :||          : | |  :  | |   :  :      || :   |  |::||
Db       1103 KPQIPVLSPLLVKVIDEKGTVDPVYLARHCREPVKMVSVLEHARDQHIITDRTIVIDVGP 1162

Qy       2910 DNTLTTLTHHNLPNPPTTTLTLTHPHHHPQTHL---LTNLAKT------TTTWHPHHYTH 2960
                    :  :    |    :: |    |   |   :   |||:  |        |      :
Db       1163 KALMAGMIKTTLDKDTSSAL----PTLGPSLDVWKSLTNILGTLYSRGLDINW----VAY 1214

Qy       2961 HDNQPHTHTHLDLPTY---------PFQ------HHHYWLESTQPGAGNVSAAGLDPTEH 3005
                 |:        ::||:|        |::     . |         ||    ::    |::
Db       1215 HEPFGSAKKVIELPSYGWDLKDYFIPYKGEWCLHRHEIRCSCATPGKETATSDYQLPSDE 1274

Qy       3006 PL---------------------------------LGATLELATDGGALLAGRLSLRSH 3031
                  :                               |||||  : ||      :::
Db       1275 QVAAKRPSKQDESKEAYPEIVATTTVHRVVEEKTEPLGATLVVETDISRPDVNQIA---- 1330

Qy       3032 PWLADHAVGGTVLLSGATFLELALHAGTYVGCDRVDELTLHAPLVVPVDGGVSVQVGVAA 3091
                    | | |  | : : : ::||| | |      :: |     |
Db       1331 ---QGHLVDGIPLCTPSVYADIALHVGRY----SMNRLRASHP---------------- 1366

Qy       3092 ADGEGRRLVSVYARGGSACGGGGASGGVWTCHASGVLVEAAAGGVVVD------------ 3139
                                  ||  ||          |: |   :|:|
Db       1367 ----------------------GAMDGV---------VDVA--DMVIDKALIPHGKSPQL 1393

Qy       3140 ---GLAGVWPPRGAVAVD----------VDGVRD------------------------- 3160
```

```
                    |  .|||: |                 ||  |
Db       1394 LRTTLTMTWPPKAAATTRSAKIKFATYFADGKLDTEHATCTVRFTSEAQLKSLQKKVPEY 1453

Qy       3161 ----RLAGAGCVLGPVFSGLRAVWRDGGDLLAEVC------------LPEEAWGDAAG-- 3202
                :  | |   |     :|   : |  |:: :              :  ||  :|
Db       1454 QERIKKLGEGLRQGQF---IRYTTKSGYKLMSSMASFHRDYKLLNHLILNEADNEAVSTM 1510

Qy       3203 ----------FGLHPALLDGVVQPLSVLLPGGTGF--GEGAGFGEGVRVPAVWGGVSLHR 3250
                      |  ||| :| : |        ||      :     : | |    |    :::
Db       1511 DFSAAKSEGTFAAHPAYVDAITQ------VGGFAMNANDNTDIQQEVFVNHGWTSFQVYQ 1564

Qy       3251 AGVTG----VRVRVSAVGRGGGREAVSVVVGDEAGVP--------------------VA 3285
                | |    | ||::   :|      ::|:  :| |                      |
Db       1565 PLVKGKTYEVYVRMTEDEKGDLVHGDTIVLYGDAVVAFFKGLSVSLSHLSETQLTSWYVR 1624

Qy       3286 SVDRLELRPVDM----------GQLRAVSVSAGRRGSL-------------YAVQWAEVG 3322
                | |  || |             |  :||   | :| :|               | ::
Db       1625 RVPRRGLRMVLQQASDKAARLHGNQQAVKTQAPQRAALKQKPQSSPTQPHASKVAYSRSA 1684

Qy       3323 PVPVCGQAWAWHEDVGESGGGPVPGVVVLRCPDAGAGGG----GGGGGGGVGEVVGGVL 3378
                |  |:       |:  |       |: :   ::|   |           |:  :   |:
Db       1685 TSPTAGKPVVAARDLSREGDDKFKAVLSVISEESGVALGELTADTNFADIGIDSLSSMVI 1744

Qy       3379 GV-VQGWLGLERFAGSRLVV----------VTRGAVV-----------------AGPEDG 3410
                |  ::  |||| |   | :          :  |: |                 | ||
Db       1745 GSRLREDLGLELGAEFSLFIDCPTVRSLKTLLSGSAVSVNNDKDELEPGQEAETAAPEQ- 1803

Qy       3411 PVDV-VGASVWGLVRSAQAEHPD------RFVLLDLDTDTGTDLD 3448
                :|: :| :    || | |  |      | ||  :  ::|  ||
Db       1804 -LDLRIGDAAPSKVRDANIEPLDLGDELFRNVLRIVSEESGVALD 1847


RESULT 15
FAS_CHICK
ID   FAS_CHICK      STANDARD;      PRT;  2511 AA.
AC   P12276;
DT   01-OCT-1989 (Rel. 12, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Fatty acid synthase (EC 2.3.1.85) [Includes: EC 2.3.1.38; EC 2.3.1.39;
DE   EC 2.3.1.41; EC 1.1.1.100; EC 4.2.1.61; EC 1.3.1.10; EC 3.1.2.14].
GN   FASN OR FAS.
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC   Gallus.
OX   NCBI_TaxID=9031;
RN   [1]
RP   SEQUENCE FROM N.A., AND SEQUENCE OF 1-11.
RC   STRAIN=White leghorn; TISSUE=Liver;
RX   MEDLINE=95031085; PubMed=7944406;
RA   Huang W.-Y., Chirala S.S., Wakil S.J.;
RT   "Amino-terminal blocking group and sequence of the animal fatty acid
RT   synthase.";
RL   Arch. Biochem. Biophys. 314:45-49(1994).
RN   [2]
RP   SEQUENCE OF 75-1775 FROM N.A.
RC   TISSUE=Liver;
RX   MEDLINE=89282777; PubMed=2734291;
RA   Holzer K.P., Liu W., Hammes G.G.;
RT   "Molecular cloning and sequencing of chicken liver fatty acid
RT   synthase cDNA.";
RL   Proc. Natl. Acad. Sci. U.S.A. 86:4387-4391(1989).
RN   [3]
RP   SEQUENCE OF 1568-2512 FROM N.A., AND PARTIAL SEQUENCE.
RX   MEDLINE=89139426; PubMed=2917973;
RA   Chirala S.S., Kasturi R., Pazirandeh M., Stolow D.T., Huang W.-Y.,
RA   Wakil S.J.;
RT   "A novel cDNA extension procedure. Isolation of chicken fatty acid
RT   synthase cDNA clones.";
RL   J. Biol. Chem. 264:3750-3757(1989).
RN   [4]
RP   SEQUENCE OF 1752-2512 FROM N.A.
RX   MEDLINE=88320436; PubMed=2842766;
RA   Yuan Z., Liu W., Hammes G.G.;
RT   "Molecular cloning and sequencing of DNA complementary to chicken
RT   liver fatty acid synthase mRNA.";
RL   Proc. Natl. Acad. Sci. U.S.A. 85:6328-6331(1988).
RN   [5]
RP   SEQUENCE OF 2202-2512 FROM N.A.
RX   MEDLINE=89088152; PubMed=3207710;
RA   Kasturi R., Chirala S.S., Pazirandeh M., Wakil S.J.;
RT   "Characterization of a genomic and cDNA clone coding for the
RT   thioesterase domain and 3' noncoding region of the chicken liver
RT   fatty acid synthase gene.";
RL   Biochemistry 27:7778-7785(1988).
RN   [6]
RP   SEQUENCE OF 2121-2209.
RX   MEDLINE=89192401; PubMed=2648999;
RA   Huang W.-Y., Stoops J.K., Wakil S.J.;
RT   "Complete amino acid sequence of chicken liver acyl carrier protein
RT   derived from the fatty acid synthase.";
RL   Arch. Biochem. Biophys. 270:92-98(1989).
RN   [7]
RP   SEQUENCE OF 2209-2508.
RC   STRAIN=White leghorn;
RX   MEDLINE=89088151; PubMed=3207709;
RA   Yang C.-Y., Huang W.-Y., Chirala S.S., Wakil S.J.;
RT   "Complete amino acid sequence of the thioesterase domain of chicken
RT   liver fatty acid synthase.";
RL   Biochemistry 27:7773-7777(1988).
RN   [8]
RP   SEQUENCE OF 667-674 AND 1698-1709.
RX   MEDLINE=89323081; PubMed=2751995;
RA   Chang S.I., Hammes G.G.;
RT   "Amino acid sequences of pyridoxal 5'-phosphate binding sites and
RT   fluorescence resonance energy transfer in chicken liver fatty acid
RT   synthase.";
RL   Biochemistry 28:3781-3788(1989).
CC   -!- FUNCTION: FATTY ACID SYNTHETASE CATALYZES THE FORMATION OF
CC       LONG-CHAIN FATTY ACIDS FROM ACETYL-COA, MALONYL-COA AND NADPH.
CC       THIS MULTIFUNCTIONAL PROTEIN HAS 7 CATALYTIC ACTIVITIES AND AN
CC       ACYL CARRIER PROTEIN.
CC   -!- CATALYTIC ACTIVITY: Acetyl-CoA + N malonyl-CoA + 2N NADPH = a
CC       long-chain fatty acid + (N+1) CoA + N CO(2) + 2N NADP(+).
CC   -!- CATALYTIC ACTIVITY: Acetyl-CoA + [acyl-carrier protein] = CoA +
CC       acetyl-[acyl-carrier protein].
CC   -!- CATALYTIC ACTIVITY: Malonyl-CoA + [acyl-carrier protein] = CoA +
CC       malonyl-[acyl-carrier protein].
CC   -!- CATALYTIC ACTIVITY: Acyl-[acyl-carrier protein] + malonyl-[acyl-
CC       carrier protein] = 3-oxoacyl-[acyl-carrier protein] + CO(2) +
CC       [acyl-carrier protein].
CC   -!- CATALYTIC ACTIVITY: (3R)-3-hydroxyacyl-[acyl-carrier protein] +
CC       NADP(+) = 3-oxoacyl-[acyl-carrier protein] + NADPH.
CC   -!- CATALYTIC ACTIVITY: (3R)-3-hydroxypalmitoyl-[acyl-carrier protein]
CC       = 2-hexadecenoyl-[acyl-carrier protein] + H(2)O.
CC   -!- CATALYTIC ACTIVITY: Acyl-[acyl-carrier protein] + NADP(+) = trans-
CC       2,3-dehydroacyl-[acyl-carrier protein] + NADPH.
CC   -!- CATALYTIC ACTIVITY: Oleoyl-[acyl-carrier protein] + H(2)O = [acyl-
CC       carrier protein] + oleate.
CC   -!- SUBUNIT: HOMODIMER, WHICH IS ARRANGED IN A HEAD TO TAIL FASHION.
CC   -!- CAUTION: REF.3 AND REF.5 SEQUENCES DIFFER FROM THAT SHOWN FROM
CC       POSITION 2351 ONWARDS DUE TO A FRAMESHIFT.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; J04485; AAB46389.1; -.
DR   EMBL; J03860; AAA48767.1; -.
DR   EMBL; J02839; AAA82106.1; ALT_SEQ.
DR   PIR; A33918; XYCHFA.
```

```
DR   PIR; A32015; A32015.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR002085; Adh_zn_family.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR003880; Ppantne_attach.
DR   InterPro; IPR001031; Thioesterase.
DR   Pfam; PF00107; adh_zinc; 1.
DR   Pfam; PF00109; ketoacyl-synt; 1.
DR   Pfam; PF00550; pp-binding; 1.
DR   Pfam; PF00698; Acyl_transf; 1.
DR   Pfam; PF00975; Thioesterase; 1.
DR   Pfam; PF02801; ketoacyl-synt_C; 1.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; 1.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 1.
DR   PROSITE; PS50075; ACP_DOMAIN; 1.
KW   Fatty acid biosynthesis; Multifunctional enzyme; Phosphopantetheine;
KW   Transferase; Hydrolase; Oxidoreductase; Lyase; NADP; Acetylation;
KW   Alternative splicing; Pyridoxal phosphate.
FT   INIT_MET      0      0
FT   DOMAIN        1   ?411       BETA-KETOACYL SYNTHASE.
FT   DOMAIN      427    815       ACYL AND MALONYL TRANSFERASES.
FT   DOMAIN     1638   1866       ENOYL REDUCTASE.
FT   DOMAIN     1867   2119       BETA-KETOACYL REDUCTASE.
FT   DOMAIN     2124   2180       ACYL CARRIER (ACP).
FT   DOMAIN     2209   2511       THIOESTERASE.
FT   MOD_RES       1      1       ACETYLATION.
FT   ACT_SITE    160    160       BETA-KETOACYL SYNTHASE (BY SIMILARITY).
FT   ACT_SITE    579    579       ACYL/MALONYL TRANSFERASES (BY
FT                                SIMILARITY).
FT   ACT_SITE    877    877       BETA-HYDROXYACYL DEHYDRATASE (BY
FT                                SIMILARITY).
FT   NP_BIND    1674   1691       NADP (ER).
FT   BINDING    1707   1707       PYRIDOXAL PHOSPHATE (BY SIMILARITY).
FT   NP_BIND    1888   1903       NADP (KR).
FT   BINDING    2157   2157       PHOSPHOPANTETHEINE (BY SIMILARITY).
FT   ACT_SITE   2308   2308       THIOESTERASE (BY SIMILARITY).
FT   ACT_SITE   2481   2481       THIOESTERASE (BY SIMILARITY).
FT   VARSPLIC   2348   2348       T -> TQCFSFSLF (IN ISOFORM 1).
FT   CONFLICT     77     78       QL -> PV (IN REF. 2).
FT   CONFLICT    116    116       L -> A (IN REF. 2).
FT   CONFLICT    675    675       R -> S (IN REF. 2).
FT   CONFLICT   1169   1169       K -> N (IN REF. 2).
FT   CONFLICT   1178   1178       A -> T (IN REF. 2).
FT   CONFLICT   1191   1191       R -> H (IN REF. 2).
FT   CONFLICT   1198   1198       P -> L (IN REF. 2).
FT   CONFLICT   1286   1287       DN -> ND (IN REF. 2).
FT   CONFLICT   1372   1372       K -> E (IN REF. 2).
FT   CONFLICT   1533   1533       C -> Y (IN REF. 2).
FT   CONFLICT   1577   1577       W -> R (IN REF. 2).
FT   CONFLICT   1685   1696       QAAIAIALSMGC -> ASSHCHRLEHGLA
FT                                (IN REF. 2; AAA48767).
FT   CONFLICT   1732   1732       Q -> E (IN REF. 2).
FT   CONFLICT   1745   1745       S -> N (IN REF. 2).
SQ   SEQUENCE   2511 AA;  274648 MW;  622039DAC8315D3F CRC64;

  Query Match             4.8%;  Score 1010;  DB 1;  Length 2511;
  Best Local Similarity   21.7%;  Pred. No. 2.7e-38;
  Matches  514;  Conservative  281;  Mismatches  782;  Indels  792;  Gaps   86;

Qy       2086 EPIAIVGMACRFPGGVTSADDFWDLISSEQDAIGGFPTDRGWDLDTLYDPDPDHPGTC-Y 2144
Db          1 EDVVIAGIAGKLPES-ENLQEFWENLLNGVDMV--TEDDRRW-----------KPGIYGL 46

Qy       2145 TRNGGFLYDAGHFDAEFFGISPREALAMDPQQRLLLETAWETIEHAGINPHTLHGTPTGV 2204
Db         47 PKRNGKLKDIKKFDASFFGVHPKQAHTMDPQLRLLLEVSYEAILDGGINPTALRGTDTGV 106

Qy       2205 FTGTNGQDYALRVHNAGQSTDGFALTGTAGSVISGRISYTFGFEGPAVSVDTACSSSLVA 2264
Db        107 WVGASGSEALEALSQDPEELLGYSMTGCQRAMLANRISYFYDFTGPSLTIDTACSSSLMA 166

Qy       2265 LHLACQALRAGECSMALAGGVTVMSSPGAFVEFSRQRGLAADGHCKAFSAAADGTGWGEG 2324
```

```
Db        167 LENAYKAIRHGQCSAALVGGVNILLKPNTSVQFMKLGMLSPDGACKAFDVSGNGYCRSEA 226

Qy       2325 VGMLLVERLSDAHRNGHRVLAVVRGSAVNQDG-ASNGLTAPNGPSQQRVIRQALANAGLS 2383
Db        227 VVVVLLTKKSMA----KRVYATIVNAGSNTDGFKEQGVTFPSGEMQQQLVGSLYRECGIK 282

Qy       2384 AGDVDAVEAHGTGTTLGDPIEAQALLATYGQDRAGEGPLWLGSVKSNVGHTQAAAGVAGV 2443
Db        283 PGDVEYVEAHGTGTKVGDPQEVNGIVNVFCQ--CEREPLLIGSTKSNMGHPEPASGLAAL 340

Qy       2444 IKMVMALRHGLLPRTLHVDEPSPHVD-WSAGAVQLLTETVPWPGGEGRLRRAGVSSFGVS 2502
Db        341 AKVILSLEHGLWAPNLHFNDPNPDIPALHDGSLKVVCKPTPVKGG-----LVSINSFGFG 395

Qy       2503 GTNAHVILEEAPADDVPGGPPAGEGDAGSDDEAAAGSPGVWPWLVSAKSQPALRAQAQAL 2562
Db        396 GSNAHVILR-------------PNEKKCQPQETCNLPRLVQ--VCGRTQEAVEILIEES 439

Qy       2563 HAHLTDHPGLDLADVGYTLAHARAV----FDHRA-TLIAADRD-TFLQALQALAAGEPHP 2616
Db        440 RKHGGCSPFLSL------LSDISAVPVSSMPYRGYTLVGTESDITEIQQVQ--ASGRP-- 489

Qy       2617 AVIHSSAPGGTGTGEAAGKTAFICSGQGTQRPGMAHGLYHTHPVFAAALNDICTHLDPHL 2676
Db        490 -------------------LWYICSGMGTQWKGMGLSLMKLDLFRQSILRS-----DEAL 525

Qy       2677 DHPLLP----LLTQNDNDNEDAAALLQQTRYAQPALFAFQVALHRLLTDGYHITPHYYAG 2732
Db        526 KSTGLKVSDLLLNADENTFDD-------TVHAFVGLAAIQIAQIDVL-KAAGLQPDGILG 577

Qy       2733 HSLGEITAAHLAGILTLTDATTLITQRATLMQ--TMPPGTMTTLHTTPHHITHHLTAHEN 2790
Db        578 HSVGELACGYADNSLSHEEAVLAAYWRGRCVKEAKLPPGGMAAVGLTWEECKQRCPP---- 634

Qy       2791 DLAIAAINTPTSLVISGTPHTVQHITTLCQQQGIKTKTL-PTNHAFHSPHTNPILNQLHQ 2849
Db        635 NVVPACHNSEDTVTVSGPLDSVSEFVTKLKKDGVFAKEVRRAGVAFHSYYMASIAPALLS 694

Qy       2850 HTQTLTYHP-PHTP-LITANTPPDQLLTPHYWTQQARN-TVDYAT---TTQTLHQHGV-- 2901
Db        695 ALKKVIPHPKPRSARWISTSIPESQ-----WQSDLARNSSAEYHVNNLVNPVLFHEGLKH 749

Qy       2902 ----TTYIELGPDNTLTTLTHHNLPNPPTTTL--TLTHPHHHPQTHLLTNLAKTTTT--- 2952
Db        750 IPENAVVVEIAPHALLQAILRRTL--KPTCTILPLMKKDHKNNLEFFLTQTGKIHLTGIN 807

Qy       2953 ------WHPHHYTHHDNQPHTHTHLDLPTYPFQHHHYW-------LESTQPGAGNVSAAG 2999
Db        808 VLGNNLFPPVEYPVPVGTP-----LISPYIKWDHSQDWDVPKAEDFPSGSKGSASASVYN 862

Qy       3000 LDPT----EHPLLGATLELATDGGALLAGRLSLRSHPWLADHAVGGTVLLSGATFLELAL 3055
Db        863 IDVSPDSPDHYLVG--------------------------HCIDGRVLYPATGYLVLAW 895

Qy       3056 HA-----GTYVGCDRV--DELTLHAPLVVPVDGGVSVQVGVAAA------DGEGRRLVS- 3101
Db        896 RTLARSLGMVMEQTAVMFEEVTIHQATILPKKGSTQLEVRIMPASHSFEVSGNGNLAVSG 955

Qy       3102 ------------------------------------VYA-------------RGGSACGG 3112
Db        956 KISLLENDALKNFHNQLADFQSQANVTAKSGLLMEDVYQELHLRGYNYGPTFQGVLECNS 1015

Qy       3113 GGASG-----GVWTCHASGVL-----------------------------------V 3129
Db       1016 EGSAGKILWNGNWVTFLDTLLHLIVLAETGRSLRLPTRIRSVYIDPVLHQEQVYQYQDNV 1075

Qy       3130 EA------------AAGGVVVDGL-AGVWPPR---------------------------- 3148
Db       1076 EAFDVVVDRCLDSLKAGGVQINGLHASVAPRRQQERISPTLEKFSFVPYIESDCLSSSTQ 1135

Qy       3149 ------------------------------------GAVA-----------------VDVD 3156
Db       1136 LHAYLEHCKGLIQKLQAKMALHGVKLVIHGLETKGAAAGSPPAQKGLQHILTEICRLELN 1195
```

```
Qy       3157 G-------------------------------------------------VRDRLAGAG- 3166
                                                                : : |||:|
Db       1196 GNPHSELEQIVTQEKMHLQDDPLLNGLLDSSELKTCLDVAKENTTSHRMKIVEALAGSGR 1255

Qy       3167 ------------------------------------------------------------ 3166

Db       1256 LFSRVQSILNTQPLLQLDYIATDCTPETLSDNETELHDAGISFSQWDPSSLPSGNLTNAD 1315

Qy       3167 --------CVLG---PVFSGLRAVWRDGGDLLAEVCLPEEAWGDAAGF------------ 3203
                      |||    : | | |  ::|| :|    | ||  |:   |
Db       1316 LAVCNCSTSVLGNTAEIISNLAAAVKEGGFVLLHTLLKEETLGEIVSFLTSPDLQQKHSF 1375

Qy       3204 ------------------------------------------------------GLHPA 3208
                                                                    |
Db       1376 LSQAQWEELFSKASLNLVAMKRSFFGSVIFLCRRQSPAKAPILLPVDDTHYKWVDSLKEI 1435

Qy       3209 LLDGVVQPLSVLLPGGTGFGEGAGFGEGVRVPAVWGGVSLHRAGVTGVRVRVSAVGRGGG 3268
              | |   ||| :    |  |           : | |:  |    | |:|   |
Db       1436 LADSSEQPLWLT---ATNCGNS----------GILGMVNCLRLEAEGHRIRCVFVSN--- 1479

Qy       3269 REAVSVVVGDEAGVPVASVDRLELR----------------------PVDMGQLRAVS- 3304
                    :   : ||  |:  ||::                       |:   | : ::
Db       1480 -------LSPSSTVPATSLSSLEMQKIIERDLVMNVYRDGKWGSFRHLPLQQAQPQELTE 1532

Qy       3305 ---VSAGRRGSLYAVQWAEVGP----------VPVCGQAWA----WHEDV----GESGGG 3343
                 |:   || | :::|  | |           | :|   :|    |  |:    |:
Db       1533 CAYVNVLTRGDLSSLRWI-VSPLRHFQTTNPNVQLCKVYYASLNFW--DIMLATGKLSPD 1589

Qy       3344 PVPGVVVLRCPDAGAGGGGGGGGGVGEVVG-----GVLGVV-----------QGWLGL 3387
               :||   |:    |    |    |    |:|     |:  ||           : |  |
Db       1590 AIPGNWTLQQCMLGMEFSGRDLAG---RRVMGLLPAKGLATVVDCDKRFLWEVPENWT-L 1645

Qy       3388 ER-------FAGSRLVVVTRGAVVAGPE------DGPVD--------VVGASVWGLVRSA 3426
                  :| :   :| || :  |          | |           :|  |:  | ||
Db       1646 EEAASVPVVYATAYYALVVRGGMKKGESVLIHSGSGGVGQAAIAIALSMGCRVFATVGSA 1705

Qy       3427 QAEH--PDRFVLLDLDTDTGTDLDT----------GAGAGWGVDG---GRVAAVVAC--- 3468
                  :       ||  || ::   :   |          | |    ::     :: | : |
Db       1706 EKREYLQARFPQLDANSFASSRNTTFQQHILRVTNGKGVSLVLNSLAEEKLQASLRCLAQ 1765

Qy       3469 -------GEPQLAVRGERLLAARLKRLESSG-------------DVPAQ------RSGD 3501
                     |:  |:   :  :|  || :   |             :| ::       : |
Db       1766 HGRFLEIGKFDLSNNSQLGMALFLKNVAFHGILLDSIFEEGNQEWEVVSELLTKGIKDGV 1825

Qy       3502 TRARRSDV-------------PAQRSGGV----------PAR-RSVDVSGREVLPWLS 3535
                :  |: |                 :  | |          | |   | :|
Db       1826 VKPLRTTVFGKEEVEAAFRFMAQGKHIGKVMIKIQEEEKQYPLRSEPVKLSAISRTSCPP 1885

Qy       3536 GGSVLVTGGTGVLGAAVARHLAGVCGVRDLLLVSRRGPDAPGAEGLRAELAALGAEVRIV 3595
                | ::||| |  |  :|: |    | : |:| || |           |  ||| :| :
Db       1886 TKSYIITGGLGGFGLELAQWLI-ERGAQKLVLTSRSGIRTGYQAKCVREWKALGIQVLVS 1944

Qy       3596 ACDVGERREVVRLLEGVPAGCPLTGVVHAAGVLDDATIASLTPERLGTVFAAKVDAALLL 3655
                 |||      |:|      |: |: : | || || | : |||    |   |    | |
Db       1945 TSDVGTLEGTQLLIEEALKLGPVGGIFNLAVVLKDAMIENQTPELFWEVNKPKYSGTLHL 2004

Qy       3656 DELTRGM--ELSAFVLFSSAAGILGSAGQGNYAAANAALDALAYRRRAAGLPGVSLAWGL 3713
                | :||    :|  ||:||| :   |:||| ||  ||:|:: :  :|   ||||::: ||
Db       2005 DWVTRKKCPDLDYFVVFSSVSCGRGNAGQSNYGFANSAMERICEQRHHDGLPGLAVQWG- 2063

Qy       3714 WEEASGMTGHLAGTDHRRIIRSGLHPMSTPDALALFDAALALDRPVLLPADLRPAPPLPP 3773
                  | |  | |    :| ::  |          | : |  :  |::|              |
Db       2064 ---AIGDVGILKAMGNREVVIGGTVLQQISSCLEVLD--MFLNQP------------HP 2105

Qy       3774 LLQDLLPATRRRTTRTTTTGGADNGAQLHARLAGQTHEQQHTTLLALVRSHIATVLGHTT 3833
              ::   : | :       : ||:                       || : :| :||
Db       2106 VMSSFVLAEK---VSVKSEGGSQRD--------------------LVEA-VAHILGVRD 2140

Qy       3834 PDTIPPDRAFRDLGFDSLTAVELRNRLSR 3862
                ::  : :  ||| |||  ||:|  | |
Db       2141 VSSLNAESSLADLGLDSLMGVEVRQTLER 2169
```

Search completed: June 17, 2003, 13:02:29
Job time : 116.316 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 17, 2003, 12:52:22 ; Search time 0.894236 Seconds
(without alignments)
3803.313 Million cell updates/sec

Title: US-09-914-286-3_COPY_3805_3886
Perfect score: 429
Sequence: 1 LAGQTHEQQHTTLLALVRSH.....RLPTTLAFDHPNPTTLTHHL 82

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	235	54.8	3567	1	ERY2_SACER	Q03132 saccharopol
2	229	53.4	3519	1	OL56_STRAT	Q07017 streptomyce
3	223	52.0	3491	1	ERY1_SACER	Q03131 saccharopol
4	207	48.3	3172	1	ERY3_SACER	Q03133 saccharopol
5	114	26.6	1538	1	PPSB_MYCTU	Q10978 mycobacteri
6	95	22.1	1774	1	MSAS_PENPA	P22367 penicillium
7	90	21.0	1876	1	PPSA_MYCTU	Q10977 mycobacteri
8	86.5	20.2	4427	1	PKSL_BACSU	Q05470 bacillus su
9	76	17.7	2110	1	MCAS_MYCBO	Q02251 mycobacteri
10	74.5	17.4	2181	1	STCA_EMENI	Q12397 emericella
11	73.5	17.1	1392	1	LYS2_YEAST	P07702 saccharomyc
12	71	16.6	3587	1	SRF1_BACSU	P27206 bacillus su
13	69.5	16.2	77	1	ACP_SYNY3	P20804 synechocyst
14	68.5	16.0	813	1	KRAF_CAEEL	Q07292 caenorhabdi
15	68.5	16.0	2505	1	FAS_RAT	P12785 rattus norv
16	68	15.9	1278	1	DHBF_BACSU	P45745 bacillus su
17	68	15.9	2035	1	HMP2_YEREN	P48633 yersinia en
18	67.5	15.7	838	1	FAS_MOUSE	P19096 mus musculu
19	67.5	15.7	4273	1	PKSM_BACSU	P40872 bacillus su
20	65.5	15.3	237	1	NOLA_BRAJA	P22537 bradyrhizob
21	64.5	15.0	76	1	ACP_ANAVA	P20803 anabaena va
22	64.5	15.0	83	1	ACP_ANASP	P58553 anabaena sp
23	64	14.9	86	1	ACPX_STRCO	Q02054 streptomyce
24	64	14.9	2511	1	FAS_CHICK	P12276 gallus gall
25	64	14.9	5255	1	BACA_BACLI	O68006 b bacitraci
26	63.5	14.8	77	1	ACP_COMTE	P80918 comamonas t
27	63.5	14.8	244	1	RIB7_YEAST	P33312 saccharomyc
28	63.5	14.8	2095	1	RRPL_TOSV	P37800 toscana vir
29	63	14.7	64	1	FAS_RABIT	P07855 oryctolagus
30	63	14.7	75	1	ACP_FUSNN	Q8rgx5 fusobacteri
31	63	14.7	287	1	VBL1_ICMV	Q08594 indian cass
32	63	14.7	1283	1	PEX1_HUMAN	O43933 homo sapien
33	63	14.7	4447	1	PKSK_BACSU	P40803 bacillus su
34	62.5	14.6	1432	1	WRN_HUMAN	Q14191 homo sapien
35	62	14.5	78	1	ACP_NEIMA	Q9jr72 neisseria m
36	62	14.5	83	1	ACPX_STRGA	P12884 streptomyce
37	62	14.5	6486	1	TYCC_BACBR	O30409 b tyrocidin
38	61.5	14.3	5217	1	HTS1_COCCA	Q01886 cochliobolu
39	61	14.2	85	1	ACPX_STRVN	P12885 streptomyce
40	61	14.2	87	1	ACPY_STRAW	Q93hc3 streptomyce
41	61	14.2	93	1	ACPA_BRUME	Q8ygp7 brucella me
42	61	14.2	115	1	ACPM_MYCLE	O69475 mycobacteri
43	61	14.2	115	1	ACPM_MYCTU	Q10500 mycobacteri
44	61	14.2	183	1	Y487_MYCTU	Q11153 mycobacteri
45	61	14.2	242	1	YHIW_ECOLI	P37638 escherichia

ALIGNMENTS

RESULT 1
ERY2_SACER
ID ERY2_SACER STANDARD; PRT; 3567 AA.
AC Q03132; Q54096;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Erythronolide synthase, modules 3 and 4 (EC 2.3.1.94) (ORF 2) (6-
DE deoxyerythronolide B synthase II) (DEBS 2).
GN ERYA.
OS Saccharopolyspora erythraea (Streptomyces erythraeus).
OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC Actinomycetales; Pseudonocardineae; Pseudonocardiaceae;
OC Saccharopolyspora.
OX NCBI_TaxID=1836;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91220065; PubMed=2024119;
RA Donadio S., Staver M.J., McAlpine J.B., Swanson S.J., Katz L.;
RT "Modular organization of genes required for complex polyketide
RT biosynthesis.";
RL Science 252:675-679(1991).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=NRRL 2338;
RX MEDLINE=92155230; PubMed=1740151;
RA Bevitt D.J., Cortes J., Haydock S.F., Leadlay P.F.;
RT "6-Deoxyerythronolide-B synthase 2 from Saccharopolyspora erythraea.
RT Cloning of the structural gene, sequence analysis and inferred domain
RT structure of the multifunctional enzyme.";
RL Eur. J. Biochem. 204:39-49(1992).
CC -!- CATALYTIC ACTIVITY: 6 malonyl-CoA + propionyl-CoA = 7 CoA + 6-
CC deoxyerythronolide B.
CC -!- COFACTOR: NADP; CONTAINS 2 COVALENTLY BOUND PHOSPHOPANTETHEINES.
CC -!- PATHWAY: COMPLEX POLYKETIDE FORMATION IN ERYTHROMYCIN
CC BIOSYNTHESIS.
CC -!- MISCELLANEOUS: IN EACH ORF OF ERYA TWO MODULES ARE PRESENT EACH
CC ENCODING FOR A FUNCTIONAL SYNTHASE SUBUNIT. THUS ERYA SHOWING 3
CC ORFS CODES FOR 6 SYNTHASE SUBUNITS. IT IS SUPPOSED THAT EACH
CC SYNTHASE PARTICIPATES IN ONE OF THE SIX FAS-LIKE ELONGATION STEPS
CC REQUIRED FOR FORMATION OF THE POLYKETIDE. MODULE 1, 2, 3, 4, 5,
CC AND 6 PARTICIPATING IN BIOSYNTHESIS STEPS 1, 2, 3, 4, 5, AND 6,
CC RESPECTIVELY.
CC -!- MISCELLANEOUS: BIOSYNTHESIS OF POLYKETIDES; ACYLTRANSFERASE (AT),
CC BETA-KETOACYL CARRIER PROTEIN SYNTHASE (KS), AND ACYL CARRIER
CC PROTEIN (ACP) FOR CHAIN ELONGATION. BETA-KETOREDUCTASE (KR),
CC DEHYDRATASE (DH), AND ENOYL REDUCTASE (ER) FOR PROCESSING OF THE
CC BETA CARBON, AND THIOESTERASE (TE) FOR RELEASE AND LACTONIZATION
CC OF THE FULL-LENGTH CHAIN.
CC -!- SIMILARITY: TO FATTY ACID SYNTHASE (FAS).
CC -!- SIMILARITY: CONTAINS 2 ACYL CARRIER DOMAINS.
CC ----------
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way

```
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; M63677; AAA26494.1; -.
DR   EMBL; X62569; CAA44448.1; -.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR002085; Adh_zn_family.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR003880; Ppantne_attach.
DR   Pfam; PF00107; adh_zinc; 1.
DR   Pfam; PF00109; ketoacyl-synt; 2.
DR   Pfam; PF00550; pp-binding; 2.
DR   Pfam; PF00698; Acyl_transf; 2.
DR   Pfam; PF02801; ketoacyl-synt_C; 2.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; 2.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 2.
DR   PROSITE; PS50075; ACP_DOMAIN; 2.
KW   Transferase; Acyltransferase; Antibiotic biosynthesis; NADP; Repeat;
KW   Phosphopantetheine; Multifunctional enzyme.
FT   DOMAIN        1   1484       MODULE 3.
FT   DOMAIN     1485   3567       MODULE 4.
FT   DOMAIN       27    488       BETA-KETOACYL SYNTHASE 1.
FT   DOMAIN      559    884       ACYLTRANSFERASE (AT) 1.
FT   DOMAIN     1130   1301       BETA-KETOACYL REDUCTASE 1 (POSSIBLY
FT                                NON-FUNCTIONAL).
FT   DOMAIN     1397   1467       ACYL CARRIER (ACP) 1.
FT   DOMAIN     1485   1943       BETA-KETOACYL SYNTHASE 2.
FT   DOMAIN     2013   2336       ACYLTRANSFERASE (AT) 2.
FT   DOMAIN     2383   3066       DEHYDRATASE/ENOYLREDUCTASE (DH/ER).
FT   DOMAIN     3139   3322       BETA-KETOACYL REDUCTASE 2.
FT   DOMAIN     3415   3485       ACYL CARRIER (ACP) 2.
FT   ACT_SITE    202    202       THIOESTER BOND.
FT   ACT_SITE    651    651       ACYL-ENZYME INTERMEDIATE.
FT   BINDING    1430   1430       PHOSPHOPANTETHEINE (BY SIMILARITY).
FT   ACT_SITE   1661   1661       THIOESTER BOND.
FT   ACT_SITE   2115   2115       ACYL-ENZYME INTERMEDIATE.
FT   NP_BIND    2961   2978       NADP (ER).
FT   NP_BIND    3142   3157       NADP (KR).
FT   BINDING    3448   3448       PHOSPHOPANTETHEINE (BY SIMILARITY).
FT   CONFLICT    438    438       R -> A (IN REF. 2).
FT   CONFLICT    480    480       T -> S (IN REF. 2).
FT   CONFLICT   1241   1241       L -> F (IN REF. 2).
FT   CONFLICT   2664   2664       G -> V (IN REF. 2).
SQ   SEQUENCE   3567 AA;  374413 MW;  EE6284F4738AA0C0 CRC64;

  Query Match             54.8%;  Score 235;  DB 1;  Length 3567;
  Best Local Similarity   56.1%;  Pred. No. 2e-17;
  Matches   46;  Conservative   10;  Mismatches   26;  Indels    0;  Gaps    0;

Qy          1 LAGQTHEQQHTTLLALVRSHIATVLGHTTPDTIPPDRAFRDLGFDSLTAVELRNRLSRTT 60
              |||::   |    |  |||||  | | |: : | :|  :||:|||||||| ||||||||    |
Db       3403 LAGRSESDQVAGLAELVRSHAAAVSGYGSADQLPERKAFKDLGFDSLAAVELRNRLGTAT 3462

Qy         61 GLRLPTTLAFDHPNPTTLTHHL 82
              |:|||:|| ||||  |   :  ||
Db       3463 GVRLPSTLVFDHPTPLAVAEHL 3484


RESULT 2
OL56_STRAT
ID   OL56_STRAT     STANDARD;      PRT;  3519 AA.
AC   Q07017;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Oleandomycin polyketide synthase, modules 5 and 6.
GN   ORFB.
OS   Streptomyces antibioticus.
OC   Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC   Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=1890;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=94150470; PubMed=8107683;
RA   Swan D.G., Rodriguez A.M., Vilches C., Mendez C., Salas J.A.;
RT   "Characterisation of a Streptomyces antibioticus gene encoding a type
RT   I polyketide synthase which has an unusual coding sequence.";
RL   Mol. Gen. Genet. 242:358-362(1994).
CC   -!- FUNCTION: MAY BE INVOLVED IN THE BIOSYNTHESIS OF THE OLEANDOMYCIN
CC       LACTONE RING.
CC   -!- COFACTOR: CONTAINS 2 COVALENTLY BOUND PHOSPHOPANTETHEINES.
CC   -!- SIMILARITY: CONTAINS 2 ACYL CARRIER DOMAINS.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; L09654; AAA19695.1; -.
DR   HSSP; P25715; 1MLA.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR003880; Ppantne_attach.
DR   InterPro; IPR001031; Thioesterase.
DR   Pfam; PF00109; ketoacyl-synt; 2.
DR   Pfam; PF00550; pp-binding; 2.
DR   Pfam; PF00698; Acyl_transf; 2.
DR   Pfam; PF00975; Thioesterase; 1.
DR   Pfam; PF02801; ketoacyl-synt_C; 2.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; 2.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 2.
DR   PROSITE; PS50075; ACP_DOMAIN; 2.
KW   Transferase; Acyltransferase; Antibiotic biosynthesis; NADP;
KW   Phosphopantetheine; Multifunctional enzyme; Repeat.
FT   DOMAIN        1      ?       MODULE 5.
FT   DOMAIN        ?   3519       MODULE 6.
FT   DOMAIN       32    501       BETA-KETOACYL SYNTHASE 1.
FT   DOMAIN      569    890       ACYLTRANSFERASE (AT) 1.
FT   DOMAIN     1200   1382       BETA-KETOACYL REDUCTASE 1.
FT   DOMAIN     1487   1561       ACYL CARRIER (ACP) 1.
FT   DOMAIN     1686   2156       BETA-KETOACYL SYNTHASE 2.
FT   DOMAIN     2220   2541       ACYLTRANSFERASE (AT) 2.
FT   DOMAIN     2856   3038       BETA-KETOACYL REDUCTASE 2.
FT   DOMAIN     3141   3215       ACYL CARRIER (ACP) 2.
FT   ACT_SITE      ?      ?       ACYL-ENZYME INTERMEDIATE.
FT   BINDING       ?      ?       PHOSPHOPANTETHEINE (BY SIMILARITY).
FT   DOMAIN     3270   3519       THIOESTERASE.
FT   ACT_SITE    210    210       BETA-KETOACYL SYNTHASE.
FT   ACT_SITE    660    660       ACYL-ENZYME INTERMEDIATE.
FT   NP_BIND    1203   1249       NADP.
FT   BINDING    1524   1524       PHOSPHOPANTETHEINE (BY SIMILARITY).
FT   ACT_SITE   1859   1859       BETA-KETOACYL SYNTHASE.
FT   ACT_SITE   2311   2311       ACYL-ENZYME INTERMEDIATE.
FT   NP_BIND    2859   2905       NADP.
FT   BINDING    3178   3178       PHOSPHOPANTETHEINE (BY SIMILARITY).
SQ   SEQUENCE   3519 AA;  368561 MW;  41AE78AAAEB61F86 CRC64;

  Query Match             53.4%;  Score 229;  DB 1;  Length 3519;
  Best Local Similarity   53.7%;  Pred. No. 9e-17;
  Matches   44;  Conservative   14;  Mismatches   24;  Indels    0;  Gaps    0;

Qy          1 LAGQTHEQQHTTLLALVRSHIATVLGHTTPDTIPPDRAFRDLGFDSLTAVELRNRLSRTT 60
              |:| :  :|   |: |||:  | |||| :   :| :|||::|||||||||||||| |:  |
Db       1479 LSGLSRVRQEEELVELVRAQAAVVLGHGSAQDVPAERAFKELGFDSLTAVELRNGLAAAT 1538

Qy         61 GLRLPTTLAFDHPNPTTLTHHL 82
              |:||| |:|||||| | :   |
Db       1539 GIRLPATMAFDHPNATAIARFL 1560


RESULT 3
ERY1_SACER
```

```
ID   ERY1_SACER     STANDARD;      PRT;  3491 AA.
AC   Q03131;
DT   01-OCT-1993 (Rel. 27, Created)
DT   01-OCT-1993 (Rel. 27, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Erythronolide synthase, modules 1 and 2 (EC 2.3.1.94) (ORF 1) (6-
DE   deoxyerythronolide B synthase I) (DEBS 1).
GN   ERYA.
OS   Saccharopolyspora erythraea (Streptomyces erythraeus).
OC   Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC   Actinomycetales; Pseudonocardineae; Pseudonocardiaceae;
OC   Saccharopolyspora.
OX   NCBI_TaxID=1836;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=91220065; PubMed=2024119;
RA   Donadio S., Staver M.J., McAlpine J.B., Swanson S.J., Katz L.;
RT   "Modular organization of genes required for complex polyketide
RT   biosynthesis.";
RL   Science 252:675-679(1991).
RN   [2]
RP   SEQUENCE OF 3474-3491 FROM N.A.
RX   MEDLINE=93231529; PubMed=8386127;
RA   Donadio S., Staver M.J.;
RT   "IS1136, an insertion element in the erythromycin gene cluster of
RT   Saccharopolyspora erythraea.";
RL   Gene 126:147-151(1993).
CC   -!- CATALYTIC ACTIVITY: 6 malonyl-CoA + propionyl-CoA = 7 CoA + 6-
CC       deoxyerythronolide B.
CC   -!- COFACTOR: NADP; CONTAINS 3 COVALENTLY BOUND PHOSPHOPANTETHEINES.
CC   -!- PATHWAY: COMPLEX POLYKETIDE FORMATION IN ERYTHROMYCIN
CC       BIOSYNTHESIS.
CC   -!- MISCELLANEOUS: IN EACH ORF OF ERYA TWO MODULES ARE PRESENT EACH
CC       ENCODING FOR A FUNCTIONAL SYNTHASE SUBUNIT. THUS ERYA SHOWING 3
CC       ORFS CODES FOR 6 SYNTHASE SUBUNITS. IT IS SUPPOSED THAT EACH
CC       SYNTHASE PARTICIPATES IN ONE OF THE SIX FAS-LIKE ELONGATION STEPS
CC       REQUIRED FOR FORMATION OF THE POLYKETIDE. MODULE 1, 2, 3, 4, 5,
CC       AND 6 PARTICIPATING IN BIOSYNTHESIS STEPS 1, 2, 3, 4, 5, AND 6,
CC       RESPECTIVELY.
CC   -!- MISCELLANEOUS: BIOSYNTHESIS OF POLYKETIDES; ACYLTRANSFERASE (AT),
CC       BETA-KETOACYL CARRIER PROTEIN SYNTHASE (KS), AND ACYL CARRIER
CC       PROTEIN (ACP) FOR CHAIN ELONGATION. BETA-KETOREDUCTASE (KR),
CC       DEHYDRATASE (DH), AND ENOYL REDUCTASE (ER) FOR PROCESSING OF THE
CC       BETA CARBON, AND THIOESTERASE (TE) FOR RELEASE AND LACTONIZATION
CC       OF THE FULL-LENGTH CHAIN.
CC   -!- SIMILARITY: TO FATTY ACID SYNTHASE (FAS).
CC   -!- SIMILARITY: CONTAINS 3 ACYL CARRIER DOMAINS.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; M63676; AAA26493.2; -.
DR   EMBL; L07626; AAA26504.1; -.
DR   HSSP; P25715; 1MLA.
DR   InterPro; IPR002198; ADH_short.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR003880; Ppantne_attach.
DR   Pfam; PF00106; adh_short; 1.
DR   Pfam; PF00109; ketoacyl-synt; 2.
DR   Pfam; PF00550; pp-binding; 3.
DR   Pfam; PF00698; Acyl_transf; 3.
DR   Pfam; PF02801; ketoacyl-synt_C; 2.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; 3.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 2.
DR   PROSITE; PS50075; ACP_DOMAIN; 3.
KW   Transferase; Acyltransferase; Antibiotic biosynthesis; NADP; Repeat;
KW   Phosphopantetheine; Multifunctional enzyme.
FT   DOMAIN       1   1972       MODULE 1.
FT   DOMAIN    1979   3491       MODULE 2.
FT   DOMAIN       1    375       ACYLTRANSFERASE (AT) 1.
FT   DOMAIN     414    484       ACYL CARRIER (ACP) 1.
FT   DOMAIN     503    961       BETA-KETOACYL SYNTHASE 1.
FT   DOMAIN    1030   1356       ACYLTRANSFERASE (AT) 2.
FT   DOMAIN    1611   1794       BETA-KETOACYL REDUCTASE 1.
FT   DOMAIN    1888   1958       ACYL CARRIER (ACP) 2.
FT   DOMAIN    1979   2441       BETA-KETOACYL SYNTHASE 2.
FT   DOMAIN    2507   2854       ACYLTRANSFERASE (AT) 3.
FT   DOMAIN    3055   3237       BETA-KETOACYL REDUCTASE 2.
FT   DOMAIN    3334   3404       ACYL CARRIER (ACP) 3.
FT   ACT_SITE   145    145       ACYL-ENZYME INTERMEDIATE.
FT   BINDING    447    447       PHOSPHOPANTETHEINE (BY SIMILARITY).
FT   ACT_SITE   677    677       BETA-KETOACYL SYNTHASE.
FT   ACT_SITE  1128   1128       ACYL-ENZYME INTERMEDIATE.
FT   NP_BIND   1614   1660       NADP.
FT   BINDING   1921   1921       PHOSPHOPANTETHEINE (BY SIMILARITY).
FT   ACT_SITE  2148   2148       BETA-KETOACYL SYNTHASE.
FT   ACT_SITE  2598   2598       ACYL-ENZYME INTERMEDIATE.
FT   NP_BIND   3058   3104       NADP.
FT   BINDING   3367   3367       PHOSPHOPANTETHEINE (BY SIMILARITY).
SQ   SEQUENCE   3491 AA;  365022 MW;  682BFC32C90FA8C4 CRC64;

  Query Match             52.0%;  Score 223;  DB 1;  Length 3491;
  Best Local Similarity   53.7%;  Pred. No. 4e-16;
  Matches   44;  Conservative   11;  Mismatches   27;  Indels    0;  Gaps    0;

Qy          1 LAGQTHEQQHTTLLALVRSHIATVLGHTTPDTIPPDRAFRDLGFDSLTAVELRNRLSRTT 60
              ||      ::   |  ||||| | |||| : : :| |:|| :|| |||:|:||||||   |   .
Db       1876 LASLPAPEREEALFELVRSHAAAVLGHASAERVPADQAFAELGVDSLSALELRNRLGAAT 1935

Qy         61 GLRLPTTLAFDHPNPTTLTHHL 82
              |:|||||  ||||:  ||  ||
Db       1936 GVRLPTTTVFDHPDVRTLAAHL 1957


RESULT 4
ERY3_SACER
ID   ERY3_SACER     STANDARD;      PRT;  3172 AA.
AC   Q03133; Q54097; Q99270;
DT   01-OCT-1993 (Rel. 27, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Erythronolide synthase, modules 5 and 6 (EC 2.3.1.94) (ORF 3) (6-
DE   deoxyerythronolide B synthase III) (DEBS 3).
GN   ERYA.
OS   Saccharopolyspora erythraea (Streptomyces erythraeus).
OC   Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC   Actinomycetales; Pseudonocardineae; Pseudonocardiaceae;
OC   Saccharopolyspora.
OX   NCBI_TaxID=1836;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NRRL 2338;
RX   MEDLINE=91043075; PubMed=2234082;
RA   Cortes J., Haydock S.F., Roberts G.A., Bevitt D.J., Leadlay P.F.;
RT   "An unusually large multifunctional polypeptide in the erythromycin-
RT   producing polyketide synthase of Saccharopolyspora erythraea.";
RL   Nature 348:176-178(1990).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=91220065; PubMed=2024119;
RA   Donadio S., Staver M.J., McAlpine J.B., Swanson S.J., Katz L.;
RT   "Modular organization of genes required for complex polyketide
RT   biosynthesis.";
RL   Science 252:675-679(1991).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NRRL 2338;
RX   MEDLINE=92155230; PubMed=1740151;
RA   Bevitt D.J., Cortes J., Haydock S.F., Leadlay P.F.;
RT   "6-Deoxyerythronolide-B synthase 2 from Saccharopolyspora erythraea.
```

```
RT   Cloning of the structural gene, sequence analysis and inferred domain
RT   structure of the multifunctional enzyme.";
RL   Eur. J. Biochem. 204:39-49(1992).
CC   -!- CATALYTIC ACTIVITY: 6 malonyl-CoA + propionyl-CoA = 7 CoA + 6-
CC       deoxyerythronolide B.
CC   -!- COFACTOR: NADP; CONTAINS 2 COVALENTLY BOUND PHOSPHOPANTETHEINES.
CC   -!- PATHWAY: COMPLEX POLYKETIDE FORMATION IN ERYTHROMYCIN
CC       BIOSYNTHESIS.
CC   -!- MISCELLANEOUS: IN EACH ORF OF ERYA TWO MODULES ARE PRESENT EACH
CC       ENCODING FOR A FUNCTIONAL SYNTHASE SUBUNIT. THUS ERYA SHOWING 3
CC       ORFS CODES FOR 6 SYNTHASE SUBUNITS. IT IS SUPPOSED THAT EACH
CC       SYNTHASE PARTICIPATES IN ONE OF THE SIX FAS-LIKE ELONGATION STEPS
CC       REQUIRED FOR FORMATION OF THE POLYKETIDE. MODULE 1, 2, 3, 4, 5,
CC       AND 6 PARTICIPATING IN BIOSYNTHESIS STEPS 1, 2, 3, 4, 5, AND 6,
CC       RESPECTIVELY.
CC   -!- MISCELLANEOUS: BIOSYNTHESIS OF POLYKETIDES; ACYLTRANSFERASE (AT),
CC       BETA-KETOACYL CARRIER PROTEIN SYNTHASE (KS), AND ACYL CARRIER
CC       PROTEIN (ACP) FOR CHAIN ELONGATION. BETA-KETOREDUCTASE (KR),
CC       DEHYDRATASE (DH), AND ENOYL REDUCTASE (ER) FOR PROCESSING OF THE
CC       BETA CARBON, AND THIOESTERASE (TE) FOR RELEASE AND LACTONIZATION
CC       OF THE FULL-LENGTH CHAIN.
CC   -!- SIMILARITY: TO FATTY ACID SYNTHASE (FAS).
CC   -!- SIMILARITY: CONTAINS 2 ACYL CARRIER DOMAINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X56107; CAA39583.1; -.
DR   EMBL; M63677; AAA26495.1; -.
DR   EMBL; X62569; CAA44449.1; -.
DR   HSSP; P00101; 1CCH.
DR   InterPro; IPR002198; ADH_short.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR003880; Ppantne_attach.
DR   InterPro; IPR001031; Thioesterase.
DR   Pfam; PF00106; adh_short; 1.
DR   Pfam; PF00109; ketoacyl-synt; 2.
DR   Pfam; PF00550; pp-binding; 2.
DR   Pfam; PF00698; Acyl_transf; 2.
DR   Pfam; PF00975; Thioesterase; 1.
DR   Pfam; PF02801; ketoacyl-synt_C; 2.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; 2.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 2.
DR   PROSITE; PS50075; ACP_DOMAIN; 2.
KW   Transferase; Acyltransferase; Antibiotic biosynthesis; NADP; Repeat;
KW   Phosphopantetheine; Multifunctional enzyme.
FT   DOMAIN        1   1484       MODULE 5.
FT   DOMAIN     1485   3172       MODULE 6.
FT   DOMAIN       37    484       BETA-KETOACYL SYNTHASE 1.
FT   DOMAIN      554    878       ACYLTRANSFERASE (AT) 1.
FT   DOMAIN     1116   1298       BETA-KETOACYL REDUCTASE 1.
FT   DOMAIN     1394   1464       ACYL CARRIER (ACP) 1.
FT   DOMAIN     1488   1954       BETA-KETOACYL SYNTHASE 2.
FT   DOMAIN     2021   2335       ACYLTRANSFERASE (AT) 2.
FT   DOMAIN     2555   2735       BETA-KETOACYL REDUCTASE 2.
FT   DOMAIN     2821   2891       ACYL CARRIER (ACP) 2.
FT   DOMAIN     2926   3172       THIOESTERASE.
FT   ACT_SITE    199    199       THIOESTER BOND.
FT   ACT_SITE    643    643       ACYL-ENZYME INTERMEDIATE.
FT   NP_BIND    1118   1164       NADP.
FT   BINDING    1427   1427       PHOSPHOPANTETHEINE (BY SIMILARITY).
FT   ACT_SITE   1661   1661       THIOESTER BOND.
FT   ACT_SITE   2112   2112       ACYL-ENZYME INTERMEDIATE.
FT   NP_BIND    2557   2605       NADP.
FT   BINDING    2854   2854       PHOSPHOPANTETHEINE (BY SIMILARITY).
FT   CONFLICT    231    231       S -> SA (IN REF. 2).
FT   CONFLICT    240    240       MISSING (IN REF. 2).
FT   CONFLICT    289    289       A -> R (IN REF. 2).
FT   CONFLICT    493    493       P -> R (IN REF. 2).
FT   CONFLICT    493    517       PEPRNSLRDTGFTLATRASAMEHRA -> ASRGTRCATPVS
FT                                RWPPAAAPWEQ (IN REF. 1).
FT   CONFLICT    510    510       A -> R (IN REF. 2).
FT   CONFLICT    513    513       M -> W (IN REF. 2).
FT   CONFLICT    525    525       E -> D (IN REF. 2).
FT   CONFLICT    536    536       R -> G (IN REF. 2).
FT   CONFLICT    547    551       GPNSP -> ARTR (IN REF. 2).
FT   CONFLICT    553    553       R -> G (IN REF. 2).
FT   CONFLICT    673    673       R -> A (IN REF. 2).
FT   CONFLICT    716    716       MISSING (IN REF. 2).
FT   CONFLICT    734    736       AHK -> GIT (IN REF. 2).
FT   CONFLICT    896    896       R -> RQR (IN REF. 2).
FT   CONFLICT    896    896       R -> RELPVYPFQRQR (IN REF. 1).
FT   CONFLICT    988    994       GVAAVPH -> VSLLSRD (IN REF. 2).
FT   CONFLICT   1108   1116       RTHPLEPLA -> ARTRWSPR (IN REF. 2).
FT   CONFLICT   1124   1126       MISSING (IN REF. 1).
FT   CONFLICT   1132   1132       L -> V (IN REF. 2).
FT   CONFLICT   1192   1192       A -> R (IN REF. 2).
FT   CONFLICT   1194   1194       MISSING (IN REF. 2).
FT   CONFLICT   1277   1278       AA -> RR (IN REF. 2).
FT   CONFLICT   1385   1390       LCDGRE -> STAER (IN REF. 2).
FT   CONFLICT   1485   1485       MISSING (IN REF. 2).
FT   CONFLICT   1518   1518       G -> R (IN REF. 2).
FT   CONFLICT   1601   1601       V -> L (IN REF. 2).
FT   CONFLICT   1724   1725       LP -> FA (IN REF. 2).
FT   CONFLICT   1732   1732       Q -> L (IN REF. 2).
FT   CONFLICT   1739   1743       GPAEG -> ARRA (IN REF. 2).
FT   CONFLICT   1762   1762       T -> S (IN REF. 2).
FT   CONFLICT   2252   2252       D -> DGAD (IN REF. 2).
FT   CONFLICT   2275   2277       QSP -> AVA (IN REF. 2).
FT   CONFLICT   2408   2408       G -> GR (IN REF. 2).
FT   CONFLICT   2420   2421       LA -> S (IN REF. 2).
FT   CONFLICT   2443   2444       NA -> TH (IN REF. 2).
FT   CONFLICT   2596   2596       A -> G (IN REF. 2).
FT   CONFLICT   2609   2609       P -> A (IN REF. 2).
FT   CONFLICT   2715   2722       RRAEGRAA -> AVRKAVRR (IN REF. 1).
FT   CONFLICT   2754   2754       D -> E (IN REF. 2).
SQ   SEQUENCE   3172 AA;  331474 MW;  DBBD5094E77DDD5F CRC64;

  Query Match             48.3%;  Score 207;  DB 1;  Length 3172;
  Best Local Similarity   48.1%;  Pred. No. 2.1e-14;
  Matches   39;  Conservative   11;  Mismatches   31;  Indels    0;  Gaps    0;

Qy          2 AGQTHEQQHTTLLALVRSHIATVLGHTTPDTIPPDRAFRDLGFDSLTAVELRNRLSRTTG 61
              |     |      ||     ||:| :|||::|| :  |: | :|||||||| |||:| : ||
Db       2810 AAPAREMTSQELLEFTHSHVAAILGHSSPDAVGQDQPFTELGFDSLTAVGLRNQLQQATG 2869

Qy         62 LRLPTTLAFDHPNPTTLTHHL 82
              | || || |:||     |  |:
Db       2870 LALPATLVFEHPTVRRLADHI 2890


RESULT 5
PPSB_MYCTU
ID   PPSB_MYCTU     STANDARD;      PRT;  1538 AA.
AC   Q10978; O53234;
DT   01-OCT-1996 (Rel. 34, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Phenolpthiocerol synthesis polyketide synthase ppsB.
GN   PPSB OR RV2932 OR MT3002 OR MTCY338.21 OR MTV011.01.
OS   Mycobacterium tuberculosis.
OC   Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC   Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1773;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H37Rv;
RX   MEDLINE=98295987; PubMed=9634230;
RA   Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA   Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekaia F.,
```

```
RA   Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA   Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S., .
RA   Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA   Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA   Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA   Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT   "Deciphering the biology of Mycobacterium tuberculosis from the
RT   complete genome sequence.";
RL   Nature 393:537-544(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CDC 1551 / Oshkosh;
RA   Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA   Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA   Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA   Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA   Bishai W.;
RT   "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT   laboratory strains.";
RL   Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: POLYKETIDE SYNTHASE INVOLVED IN PHENOLPTHIOCEROL
CC       SYNTHESIS.
CC   -!- COFACTOR: CONTAINS 1 COVALENTLY BOUND PHOSPHOPANTETHEINE (By
CC       similarity).
CC   -!- SIMILARITY: CONTAINS 1 ACYL CARRIER DOMAIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AL021070; CAA15929.1; -.
DR   EMBL; AE007122; AAK47329.1; -.
DR   TIGR; MT3002; -.
DR   TubercuList; Rv2932; -.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR003880; Ppantne_attach.
DR   Pfam; PF00109; ketoacyl-synt; 1.
DR   Pfam; PF00550; pp-binding; 1.
DR   Pfam; PF00698; Acyl_transf; 1.
DR   Pfam; PF02801; ketoacyl-synt_C; 1.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 1.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; FALSE_NEG.
DR   PROSITE; PS50075; ACP_DOMAIN; 1.
KW   Hypothetical protein; Multifunctional enzyme; Oxidoreductase;
KW   Transferase; NADP; Phosphopantetheine; Complete proteome.
FT   DOMAIN     1425   1495       ACYL CARRIER (ACP).
FT   BINDING    1458   1458       PHOSPHOPANTETHEINE (POTENTIAL).
SQ   SEQUENCE   1538 AA;  162527 MW;  B55E2A2042AD00CC CRC64;

  Query Match             26.6%;  Score 114;  DB 1;  Length 1538;
  Best Local Similarity   35.5%;  Pred. No. 0.00014;
  Matches   27;  Conservative   14;  Mismatches   35;  Indels    0;  Gaps    0;

Qy          7 EQQHTTLLALVRSHIATVLGHTTPDTIPPDRAFRDLGFDSLTAVELRNRLSRTTGLRLPT 66
              |::   |   | :  |||:|    : : |   |  || ||| :| |:  || : |  ||
Db       1419 EKRRDMLFDHVGALAATVMGMPPTEPLDPSAGFFQLGMDSLMSVTLQRALSESLGEFLPA 1478

Qy         67 TLAFDHPNPTTLTHHL 82
              :: ||:|   :|| :|
Db       1479 SVVFDYPTVYSLTDYL 1494


RESULT 6
MSAS_PENPA
ID   MSAS_PENPA     STANDARD;      PRT;  1774 AA.
AC   P22367;
DT   01-AUG-1991 (Rel. 19, Created)
DT   01-AUG-1991 (Rel. 19, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   6-methylsalicylic acid synthase (EC 2.3.1.-) (MSAS).
OS   Penicillium patulum (Penicillium griseofulvum).
OC   Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC   Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Penicillium.
OX   NCBI_TaxID=5078;
RN   [1]
RP   SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC   STRAIN=DSM 62862;
RX   MEDLINE=91006137; PubMed=2209605;
RA   Beck J., Ripka S., Siegner A., Schiltz E., Schweizer E.;
RT   "The multifunctional 6-methylsalicylic acid synthase gene of
RT   Penicillium patulum. Its gene structure relative to that of other
RT   polyketide synthases.";
RL   Eur. J. Biochem. 192:487-498(1990).
CC   -!- FUNCTION: This multifunctional enzyme is a polyketide synthase.
CC       It catalyzes a total of 11 steps by seven different component
CC       enzymes, in the biosynthesis of the antibiotic patulin.
CC   -!- CATALYTIC ACTIVITY: Acetyl-CoA + 3 malonyl-CoA + NADPH =
CC       6-methylsalicylic acid + NADP(+) + 3 CoA + 3 CO(2) + H(2)O.
CC   -!- PATHWAY: Patulin biosynthesis.
CC   -!- SUBUNIT: HOMOMULTIMER.
CC   -!- INDUCTION: IN THE LATE LOGARITHMIC GROWTH PHASE.
CC   -!- SIMILARITY: WITH MOST POLYKETIDE SYNTHASES, SPECIALLY WITH RAT
CC       FATTY ACID SYNTHASE, AND WITH OTHER ENZYMES SUCH AS LIPASES AND
CC       THIOLASES.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X55776; CAA39295.1; -.
DR   PIR; S13178; S13178.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR003880; Ppantne_attach.
DR   Pfam; PF00109; ketoacyl-synt; 1.
DR   Pfam; PF00550; pp-binding; 1.
DR   Pfam; PF00698; Acyl_transf; 1.
DR   Pfam; PF02801; ketoacyl-synt_C; 1.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; 1.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 1.
DR   PROSITE; PS50075; ACP_DOMAIN; 1.
KW   Multifunctional enzyme; Oxidoreductase; Antibiotic biosynthesis;
KW   Transferase; NADP; Phosphopantetheine.
FT   DOMAIN      186    238       ACYLTRANSFERASE (AT).
FT   DOMAIN      642    676       ACETYL/MALONYL TRANSFERASES.
FT   DOMAIN     1403   1450       2-OXOACYL REDUCTASE.
FT   DOMAIN     1700   1769       ACYL CARRIER (ACP).
FT   NP_BIND    1419   1424       NADP (POTENTIAL).
FT   ACT_SITE    204    204       BETA-KETOACYL SYNTHASE (BY SIMILARITY).
FT   ACT_SITE    653    653       MALONYLTRANSFERASE (BY SIMILARITY).
FT   BINDING    1732   1732       PHOSPHOPANTETHEINE (BY SIMILARITY).
SQ   SEQUENCE   1774 AA;  190732 MW;  05ED5DD10863F938 CRC64;

  Query Match             22.1%;  Score 95;  DB 1;  Length 1774;
  Best Local Similarity   37.5%;  Pred. No. 0.02;
  Matches   24;  Conservative   10;  Mismatches   26;  Indels    4;  Gaps    2;

Qy         17 VRSHIATVLGHTTPDTIPPDRAFRDLGFDSLTAVELRNRLSRTTGLRLPTTLAFDHPNPT 76
              :|  :| ||  |  | :    |  ||| ||:  | || :|  |  : :| || : ||
Db       1704 IRGCVAKVLQMTAED-VDSKAALADLGVDSVMTVTLRRQLQLTLKIAVPPTLTWSHP--- 1759

Qy         77 TLTH 80
              |::|
Db       1760 TVSH 1763


RESULT 7
```

```
PPSA_MYCTU
ID   PPSA_MYCTU     STANDARD;      PRT;  1876 AA.
AC   Q10977;
DT   01-OCT-1996 (Rel. 34, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Phenolpthiocerol synthesis polyketide synthase ppsA.
GN   PPSA OR RV2931 OR MT3000 OR MTCY338.20.
OS   Mycobacterium tuberculosis.
OC   Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC   Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1773;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H37Rv;
RX   MEDLINE=98295987; PubMed=9634230;
RA   Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA   Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekaia F.,
RA   Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA   Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA   Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA   Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA   Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA   Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT   "Deciphering the biology of Mycobacterium tuberculosis from the
RT   complete genome sequence.";
RL   Nature 393:537-544(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CDC 1551 / Oshkosh;
RA   Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA   Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA   Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA   Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA   Bishai W.;
RT   "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT   laboratory strains.";
RL   Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: POLYKETIDE SYNTHASE INVOLVED IN PHENOLPTHIOCEROL
CC       SYNTHESIS.
CC   -!- COFACTOR: CONTAINS 2 COVALENTLY BOUND PHOSPHOPANTETHEINES (By
CC       similarity).
CC   -!- SIMILARITY: CONTAINS 2 ACYL CARRIER DOMAINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z74697; CAA98988.1; -.
DR   EMBL; AE007122; AAK47328.1; -.
DR   TIGR; MT3000; -.
DR   TubercuList; Rv2931; -.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR003880; Ppantne_attach.
DR   Pfam; PF00109; ketoacyl-synt; 1.
DR   Pfam; PF00550; pp-binding; 2.
DR   Pfam; PF00698; Acyl_transf; 1.
DR   Pfam; PF02801; ketoacyl-synt_C; 1.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; 1.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 1.
DR   PROSITE; PS50075; ACP_DOMAIN; 2.
KW   Multifunctional enzyme; Oxidoreductase; Transferase; NADP; Repeat;
KW   Phosphopantetheine; Complete proteome.
FT   DOMAIN        7     80       ACYL CARRIER (ACP) 1.
FT   DOMAIN     1764   1833       ACYL CARRIER (ACP) 2.
FT   NP_BIND    1498   1503       NADP (POTENTIAL).
FT   BINDING      62     62       PHOSPHOPANTETHEINE (POTENTIAL).
FT   ACT_SITE    273    273       BETA-KETOACYL SYNTHASE (BY SIMILARITY).
FT   ACT_SITE    720    720       MALONYLTRANSFERASE (BY SIMILARITY).
FT   BINDING    1796   1796       PHOSPHOPANTETHEINE (POTENTIAL).
FT   CONFLICT    624    624       D -> E (IN REF. 2).
FT   CONFLICT    877    877       R -> H (IN REF. 2).
FT   CONFLICT   1323   1323       G -> S (IN REF. 2).
SQ   SEQUENCE   1876 AA;  198834 MW;  D9783DBD48792110 CRC64;

  Query Match             21.0%;  Score 90;  DB 1;  Length 1876;
  Best Local Similarity   36.2%;  Pred. No. 0.076;
  Matches   17;  Conservative   10;  Mismatches   20;  Indels    0;  Gaps    0;

Qy         36 DRAFRDLGFDSLTAVELRNRLSRTTGLRLPTTLAFDHPNPTTLTHHL 82
              || | :|| :|| |: :|    :  |: |  |: |:||   :|  :|
Db       1786 DRPFAELGLNSLMAMAIRREAEQFVGIELSATMLFNHPTVKSLASYL 1832


RESULT 8
PKSL_BACSU
ID   PKSL_BACSU     STANDARD;      PRT;  4427 AA.
AC   Q05470;
DT   01-FEB-1995 (Rel. 31, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Putative polyketide synthase pksL (PKS).
GN   PKSL OR PKSX OR PKSA OR OUTG.
OS   Bacillus subtilis.
OC   Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX   NCBI_TaxID=1423;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168 / PB1424;
RX   MEDLINE=93345824; PubMed=8344529;
RA   Scotti C., Piatti M., Cuzzoni A., Perani P., Tognoni A., Grandi G.,
RA   Galizzi A., Albertini A.M.;
RT   "A Bacillus subtilis large ORF coding for a polypeptide highly
RT   similar to polyketide synthases.";
RL   Gene 130:65-71(1993).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168;
RX   MEDLINE=98044033; PubMed=9384377;
RA   Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA   Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
RA   Borriss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA   Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA   Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
RA   Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,
RA   Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA   Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA   Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
RA   Guiseppi G., Guy B.J., Haga K., Haiech J., Harwood C.R., Henaut A.,
RA   Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA   Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
RA   Kobayashi Y., Koetter P., Koningstein G., Krogh S., Kumano M.,
RA   Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA   Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,
RA   Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA   Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA   Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA   Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA   Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadaie Y.,
RA   Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA   Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,
RA   Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA   Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA   Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA   Viari A., Wambutt R., Wedler E., Wedler H., Weitzenegger T.,
RA   Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA   Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.;
RT   "The complete genome sequence of the Gram-positive bacterium Bacillus
RT   subtilis.";
RL   Nature 390:249-256(1997).
RN   [3]
```

```
RP   SEQUENCE OF 3619-4427 FROM N.A.
RC   STRAIN=168 / PB1424;
RA   Grandi G.;
RL   Submitted (JUL-1994) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: POTENTIALLY INVOLVED IN SOME INTERMEDIATE STEPS FOR
CC       THE SYNTHESIS OF A POLYKETIDE MOLECULE WHICH MAY BE INVOLVED IN
CC       SECONDARY METABOLISM.
CC   -!- COFACTOR: CONTAINS 5 COVALENTLY BOUND PHOSPHOPANTETHEINES
CC       (POTENTIAL).
CC   -!- SIMILARITY: CONTAINS 5 ACYL CARRIER DOMAINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z14098; CAA78479.1; -.
DR   EMBL; U11039; AAA85145.1; -.
DR   EMBL; Z99113; CAB13602.1; -.
DR   EMBL; Z35133; CAA84504.1; -.
DR   PIR; S25021; S25021.
DR   SubtiList; BG10698; pksL.
DR   InterPro; IPR002198; ADH_short.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR003880; Ppantne_attach.
DR   Pfam; PF00106; adh_short; 1.
DR   Pfam; PF00109; ketoacyl-synt; 4.
DR   Pfam; PF00550; pp-binding; 5.
DR   Pfam; PF02801; ketoacyl-synt_C; 4.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; 5.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 1.
DR   PROSITE; PS50075; ACP_DOMAIN; 5.
KW   Transferase; Acyltransferase; Antibiotic biosynthesis; NADP;
KW   Phosphopantetheine; Multifunctional enzyme; Repeat; Complete proteome.
FT   DOMAIN      211    280       ACYL CARRIER (ACP) 1.
FT   DOMAIN      382    759       BETA-KETOACYL SYNTHASE 1.
FT   DOMAIN      937   1115       DEHYDRATASE.
FT   DOMAIN     1409   1602       BETA-KETOACYL REDUCTASE 1.
FT   DOMAIN     1687   1759       ACYL CARRIER (ACP) 2.
FT   DOMAIN     1876   2253       BETA-KETOACYL SYNTHASE 2.
FT   DOMAIN     2491   2560       ACYL CARRIER (ACP) 3.
FT   DOMAIN     2632   2701       ACYL CARRIER (ACP) 4.
FT   DOMAIN     2823   3182       BETA-KETOACYL SYNTHASE 3.
FT   DOMAIN     3575   3776       BETA-KETOACYL REDUCTASE 2.
FT   DOMAIN     3854   3923       ACYL CARRIER (ACP) 5.
FT   DOMAIN     4019   4373       BETA-KETOACYL SYNTHASE 4.
FT   BINDING     243    243       PHOSPHOPANTETHEINE (POTENTIAL).
FT   BINDING    1723   1723       PHOSPHOPANTETHEINE (POTENTIAL).
FT   BINDING    2523   2523       PHOSPHOPANTETHEINE (POTENTIAL).
FT   BINDING    2664   2664       PHOSPHOPANTETHEINE (POTENTIAL).
FT   BINDING    3886   3886       PHOSPHOPANTETHEINE (POTENTIAL).
SQ   SEQUENCE   4427 AA;  493398 MW;  9612521E561AB9F2 CRC64;

  Query Match             20.2%;  Score 86.5;  DB 1;  Length 4427;
  Best Local Similarity   35.3%;  Pred. No. 0.48;
  Matches   24;  Conservative   11;  Mismatches   24;  Indels    9;  Gaps    3;

Qy         21 IATVLGHTTPDTIPPDRAFRDLGFDSLTAVELRNRLSRTTGLRLPTTLAFDHPNPTT--- 77
              :| || :   : | || || |:| ||:| :|    :::  |  |  | :|: |||
Db       2499 LAEVL-YMDQNEIDPDEAPIDIGMDSITGLEWIKAINKQYGTSLNVTKVYDY--PTTRDF 2555

Qy         78 ---LTHHL 82
                 | | |
Db       2556 AVYLAHEL 2563


RESULT 9
MCAS_MYCBO
ID   MCAS_MYCBO     STANDARD;      PRT;  2110 AA.
AC   Q02251;
DT   01-JUN-1994 (Rel. 29, Created)
DT   01-JUN-1994 (Rel. 29, Last sequence update)
DT   15-JUL-1998 (Rel. 36, Last annotation update)
DE   Mycocerosic acid synthase.
GN   MAS.
OS   Mycobacterium bovis.
OC   Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC   Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1765;
RN   [1]
RP   SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX   MEDLINE=92406887; PubMed=1527058;
RA   Mathur M., Kolattukudy P.E.;
RT   "Molecular cloning and sequencing of the gene for mycocerosic acid
RT   synthase, a novel fatty acid elongating multifunctional enzyme, from
RT   Mycobacterium tuberculosis var. bovis Bacillus Calmette-Guerin.";
RL   J. Biol. Chem. 267:19388-19395(1992).
CC   -!- FUNCTION: CATALYZES THE ELONGATION OF N-FATTY ACYL-COA WITH
CC       METHYLAMALONYL-COA (NOT MALONYL-COA) AS THE ELONGATING AGENT TO
CC       FORM MYCOCEROSYL LIPIDS.
CC   -!- COFACTOR: CONTAINS 1 COVALENTLY BOUND PHOSPHOPANTETHEINE.
CC   -!- SUBUNIT: HOMODIMER WHOSE MONOMERS PROBABLY HAVE A HEAD TO TAIL
CC       ARRANGEMENT.
CC   -!- SUBCELLULAR LOCATION: Membrane-associated.
CC   -!- SIMILARITY: CONTAINS 1 ACYL CARRIER DOMAIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M95808; AAA25369.1; -.
DR   PIR; B44110; B44110.
DR   HSSP; P73283; 1E5M.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR002085; Adh_zn_family.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR003880; Ppantne_attach.
DR   Pfam; PF00107; adh_zinc; 1.
DR   Pfam; PF00109; ketoacyl-synt; 1.
DR   Pfam; PF00550; pp-binding; 1.
DR   Pfam; PF00698; Acyl_transf; 1.
DR   Pfam; PF02801; ketoacyl-synt_C; 1.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; 1.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 1.
DR   PROSITE; PS50075; ACP_DOMAIN; 1.
KW   Fatty acid biosynthesis; Multifunctional enzyme; Phosphopantetheine;
KW   Transferase; Hydrolase; Oxidoreductase; Ligase; NADP; Membrane.
FT   DOMAIN        1    430       BETA-KETOACYL SYNTHASE.
FT   DOMAIN      533    852       ACYL TRANSFERASE.
FT   DOMAIN        ?      ?       ENOYL REDUCTASE.
FT   DOMAIN        ?      ?       BETA-KETOACYL REDUCTASE.
FT   DOMAIN     2026   2096       ACYL CARRIER (ACP).
FT   ACT_SITE    177    177       BETA-KETOACYL SYNTHASE (BY SIMILARITY).
FT   ACT_SITE    623    623       ACYL TRANSFERASES (BY SIMILARITY).
FT   NP_BIND    1561   1578       NADP (ER).
FT   NP_BIND    1765   1780       NADP (KR).
FT   BINDING    2059   2059       PHOSPHOPANTETHEINE (BY SIMILARITY).
SQ   SEQUENCE   2110 AA;  225577 MW;  CC658215D7155300 CRC64;

  Query Match             17.7%;  Score 76;  DB 1;  Length 2110;
  Best Local Similarity   33.3%;  Pred. No. 3;
  Matches   17;  Conservative    9;  Mismatches   25;  Indels    0;  Gaps    0;

Qy         32 TIPPDRAFRDLGFDSLTAVELRNRLSRTTGLRLPTTLAFDHPNPTTLTHHL 82
              ||  ||:| : | |||  :|:|  :   ||:||   :   :      | :|
Db       2045 TIDADRSFIEYGLDSLGMLEMRTHVETETGIRLTPKVIATNKTARALAQYL 2095


RESULT 10
```

```
STCA_EMENI
ID   STCA_EMENI     STANDARD;      PRT;  2181 AA.
AC   Q12397;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Putative sterigmatocystin biosynthesis polyketide synthase (PKS).
GN   STCA OR PKSST.
OS   Emericella nidulans (Aspergillus nidulans).
OC   Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC   Eurotiales; Trichocomaceae; Emericella.
OX   NCBI_TaxID=5072;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=FGSC 26;
RX   MEDLINE=96202293; PubMed=8643646;
RA   Brown D.W., Yu J.-H., Kelkar H.S., Fernandes M., Nesbitt T.C.,
RA   Keller N.P., Adams T.H., Leonard T.J.;
RT   "Twenty-five coregulated transcripts define a sterigmatocystin gene
RT   cluster in Aspergillus nidulans.";
RL   Proc. Natl. Acad. Sci. U.S.A. 93:1418-1422(1996).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=FGSC 4;
RX   MEDLINE=95370159; PubMed=7642507;
RA   Yu J.-H., Leonard T.J.;
RT   "Sterigmatocystin biosynthesis in Aspergillus nidulans requires a
RT   novel type I polyketide synthase.";
RL   J. Bacteriol. 177:4792-4800(1995).
CC   -!- FUNCTION: INVOLVED IN THE SYNTHESIS OF THE POLYKETIDE NUCLEUS OF
CC       STERIGMATOCYSTIN FROM HEXANOYL-COA AND SEVEN MALONATES.
CC   -!- COFACTOR: CONTAINS 2 COVALENTLY BOUND PHOSPHOPANTETHEINES
CC       (POTENTIAL).
CC   -!- PATHWAY: Sterigmatocystin biosynthesis; first step.
CC   -!- SIMILARITY: CONTAINS 2 ACYL CARRIER DOMAINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U34740; AAC49191.1; -.
DR   EMBL; L39121; AAA81586.1; -.
DR   HSSP; P25715; 1MLA.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR003880; Ppantne_attach.
DR   InterPro; IPR001031; Thioesterase.
DR   Pfam; PF00109; ketoacyl-synt; 1.
DR   Pfam; PF00550; pp-binding; 2.
DR   Pfam; PF00698; Acyl_transf; 1.
DR   Pfam; PF00975; Thioesterase; 1.
DR   Pfam; PF02801; ketoacyl-synt_C; 1.
DR   PROSITE; PS50075; ACP_DOMAIN; 2.
KW   Transferase; Acyltransferase; Phosphopantetheine; Repeat;
KW   Multifunctional enzyme.
FT   DOMAIN      383    814       BETA-KETOACYL SYNTHASE.
FT   DOMAIN      884   1209       ACYL/MALONYL TRANSFERASES.
FT   DOMAIN     1706   1777       ACYL CARRIER (ACP) 1.
FT   DOMAIN     1830   1901       ACYL CARRIER (ACP) 2.
FT   DOMAIN        ?   2181       THIOESTERASE.
FT   ACT_SITE    552    552       BETA-KETOACYL SYNTHASE (BY SIMILARITY).
FT   ACT_SITE    978    978       ACYL/MALONYL TRANSFERASES (BY
FT                                SIMILARITY).
FT   BINDING    1738   1738       PHOSPHOPANTETHEINE (BY SIMILARITY).
FT   BINDING    1862   1862       PHOSPHOPANTETHEINE (BY SIMILARITY).
FT   ACT_SITE   2028   2028       THIOESTERASE (BY SIMILARITY).
SQ   SEQUENCE   2181 AA;  238831 MW;  5A3E5712AA9AD942 CRC64;

  Query Match             17.4%;  Score 74.5;  DB 1;  Length 2181;
  Best Local Similarity   28.4%;  Pred. No. 4.5;
  Matches   21;  Conservative   14;  Mismatches   32;  Indels    7;  Gaps    2;

Qy          7 EQQHTTLLALVRSHIATVLGHTTPDTIPPDRAFRDLGFDSLTAVELRNRLSRTTGLRLPT 66
              : :   :|:::        ||  | ||       | |:| |||::: : :||    || |
Db       1704 DDKFKAVLSVISEESGVALGELTADT-----NFADIGIDSLSSMVIGSRLREDLGLELGA 1758

Qy         67 --TLAFDHPNPTTL 78
                :|  | |   :|
Db       1759 EFSLFIDCPTVRSL 1772


RESULT 11
LYS2_YEAST
ID   LYS2_YEAST     STANDARD;      PRT;  1392 AA.
AC   P07702;
DT   01-APR-1988 (Rel. 07, Created)
DT   01-MAR-1992 (Rel. 21, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Aminoadipate-semialdehyde dehydrogenase large subunit (EC 1.2.1.31)
DE   (Alpha-aminoadipate reductase) (Alpha-AR).
GN   LYS2 OR YBR115C OR YBR0910.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=91192607; PubMed=2013406;
RA   Morris M.E., Jinks-Robertson S.;
RT   "Nucleotide sequence of the LYS2 gene of Saccharomyces cerevisiae:
RT   homology to Bacillus brevis tyrocidine synthetase 1.";
RL   Gene 98:141-145(1991).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288c;
RX   MEDLINE=95208357; PubMed=7900426;
RA   Mannhaupt G., Stucka R., Ehnle S., Vetter I., Feldmann H.;
RT   "Analysis of a 70 kb region on the right arm of yeast chromosome II.";
RL   Yeast 10:1363-1381(1994).
RN   [3]
RP   SEQUENCE OF 1-150 AND 1209-1392 FROM N.A.
RX   MEDLINE=87106859; PubMed=3542721;
RA   Fleig U.N., Pridmore R.D., Philippsen P.;
RT   "Construction of LYS2 cartridges for use in genetic manipulations of
RT   Saccharomyces cerevisiae.";
RL   Gene 46:237-245(1986).
RN   [4]
RP   SEQUENCE OF 1083-1392 FROM N.A.
RC   STRAIN=S288c;
RX   MEDLINE=92327848; PubMed=1626431;
RA   Mannhaupt G., Stucka R., Ehnle S., Vetter I., Feldmann H.;
RT   "Molecular analysis of yeast chromosome II between CMD1 and LYS2: the
RT   excision repair gene RAD16 located in this region belongs to a novel
RT   group of double-finger proteins.";
RL   Yeast 8:397-408(1992).
RN   [5]
RP   SEQUENCE OF 1-130 FROM N.A.
RC   STRAIN=S288c;
RX   MEDLINE=94039074; PubMed=7916691;
RA   Schaaff-Gerstenschlager I., Mannhaupt G., Vetter I., Zimmermann F.K.,
RA   Feldmann H.;
RT   "TKL2, a second transketolase gene of Saccharomyces cerevisiae.
RT   Cloning, sequence and deletion analysis of the gene.";
RL   Eur. J. Biochem. 217:487-492(1993).
CC   -!- FUNCTION: CATALYZES THE ACTIVATION OF ALPHA-AMINOADIPATE BY
CC       ATP-DEPENDENT ADENYLATION AND THE REDUCTION OF ACTIVATED
CC       ALPHA-AMINOADIPATE BY NADPH.
CC   -!- CATALYTIC ACTIVITY: L-2-aminoadipate 6-semialdehyde + NADP(+) +
CC       H(2)O = L-2-aminoadipate + NADPH.
CC   -!- COFACTOR: CONTAINS 1 COVALENTLY BOUND PHOSPHOPANTETHEINE
CC       (POTENTIAL).
CC   -!- PATHWAY: Lysine biosynthesis; sixth step.
```

```
CC   -!- SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA SUBUNIT.
CC   -!- SIMILARITY: BELONGS TO THE ATP-DEPENDENT AMP-BINDING ENZYME
CC       FAMILY.
CC   -!- SIMILARITY: CONTAINS 1 ACYL CARRIER DOMAIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M36287; AAA34747.1; -.
DR   EMBL; X66247; CAA46975.1; -.
DR   EMBL; X78993; CAA55617.1; -.
DR   EMBL; Z35984; CAA85072.1; -.
DR   EMBL; X73532; CAA51938.1; -.
DR   PIR; JU0448; YGBYAD.
DR   HSSP; P14687; 1AMU.
DR   SGD; S0000319; LYS2.
DR   InterPro; IPR000873; AMP-bind.
DR   InterPro; IPR003880; Ppantne_attach.
DR   Pfam; PF00501; AMP-binding; 1.
DR   Pfam; PF00550; pp-binding; 1.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; 1.
DR   PROSITE; PS00455; AMP_BINDING; 1.
DR   PROSITE; PS50075; ACP_DOMAIN; 1.
KW   Lysine biosynthesis; Oxidoreductase; NADP; Phosphopantetheine.
FT   DOMAIN      848    917       ACYL CARRIER (ACP).
FT   BINDING     880    880       PHOSPHOPANTETHEINE (POTENTIAL).
SQ   SEQUENCE   1392 AA;  155345 MW;  F0083A80BC6F7FB5 CRC64;

  Query Match            17.1%;  Score 73.5;  DB 1;  Length 1392;
  Best Local Similarity  30.0%;  Pred. No. 3.5;
  Matches   21;  Conservative   14;  Mismatches   34;  Indels    1;  Gaps    1;

Qy          4 QTHEQQHTTLLALVRSHIATVLGHTTPDTIPPDRAFRDLGFDSLTAVELRNRLSRTTGLR 63
              :| : | | :   ||    ::|  | | :: || :| |||  |: | ::   | :   :
Db        839 ETDDSQFTNVEREVRDLWLSIL-PTKPASVSPDDSFFDLGGHSILATKMIFTLKKKLQVD 897

Qy         64 LPTTLAFDHP 73
              ||    | :|
Db        898 LPLGTIFKYP 907


RESULT 12
SRF1_BACSU
ID   SRF1_BACSU     STANDARD;      PRT;  3587 AA.
AC   P27206;
DT   01-AUG-1992 (Rel. 23, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Surfactin synthetase subunit 1.
GN   SRFAA OR SRFA1 OR SRFA.
OS   Bacillus subtilis.
OC   Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX   NCBI_TaxID=1423;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168;
RX   MEDLINE=93181186; PubMed=8441623;
RA   Fuma S., Fujishima Y., Corbell N., D'Souza C., Nakano M.M.,
RA   Zuber P., Yamane K.;
RT   "Nucleotide sequence of 5' portion of srfA that contains the region
RT   required for competence establishment in Bacillus subtilis.";
RL   Nucleic Acids Res. 21:93-97(1993).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168 / JH642;
RX   MEDLINE=93360813; PubMed=8355609;
RA   Cosmina P., Rodriguez F., de Ferra F., Grandi G., Perego M.,
RA   Venema G., van Sinderen D.;
RT   "Sequence and analysis of the genetic locus responsible for surfactin
RT   synthesis in Bacillus subtilis.";
RL   Mol. Microbiol. 8:821-831(1993).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168;
RX   MEDLINE=97124189; PubMed=8969502;
RA   Yamane K., Kumano M., Kurita K.;
RT   "The 25 degrees-36 degrees region of the Bacillus subtilis
RT   chromosome: determination of the sequence of a 146 kb segment and
RT   identification of 113 genes.";
RL   Microbiology 142:3047-3056(1996).
RN   [4]
RP   SEQUENCE OF 1-460 FROM N.A.
RX   MEDLINE=91154134; PubMed=1847909;
RA   Nakano M.M., Magnuson R., Myers A.M., Curry J., Grossman A.D.,
RA   Zuber P.;
RT   "srfA is an operon required for surfactin production, competence
RT   development, and efficient sporulation in Bacillus subtilis.";
RL   J. Bacteriol. 173:1770-1778(1991).
RN   [5]
RP   SEQUENCE OF 1-38 FROM N.A.
RX   MEDLINE=91358326; PubMed=1715856;
RA   Nakano M.M., Xia L., Zuber P.;
RT   "Transcription initiation region of the srfA operon, which is
RT   controlled by the comP-comA signal transduction system in Bacillus
RT   subtilis.";
RL   J. Bacteriol. 173:5487-5493(1991).
RN   [6]
RP   SEQUENCE OF 1-64 FROM N.A.
RC   STRAIN=168;
RX   MEDLINE=95219080; PubMed=7704255;
RA   Fujishima Y., Yamane K.;
RT   "A 10 kb nucleotide sequence at the 5' flanking region (32 degrees)
RT   of srfAA of the Bacillus subtilis chromosome.";
RL   Microbiology 141:277-279(1995).
CC   -!- FUNCTION: THIS PROTEIN IS A MULTIFUNCTIONAL ENZYME ABLE TO
CC       ACTIVATE AND POLYMERIZE THE AMINO ACIDS LEU, GLU, ASP AND VAL.
CC       ACTIVATION SITES FOR THESE AA CONSIST OF INDIVIDUAL DOMAINS.
CC   -!- COFACTOR: CONTAINS 3 COVALENTLY BOUND PHOSPHOPANTETHEINES.
CC   -!- PATHWAY: Cyclic peptide antibiotic surfactin biosynthesis.
CC   -!- SIMILARITY: BELONGS TO THE ATP-DEPENDENT AMP-BINDING ENZYME
CC       FAMILY.
CC   -!- SIMILARITY: CONTAINS 3 ACYL CARRIER DOMAINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; D13262; BAA02522.1; -.
DR   EMBL; X70356; CAA49816.1; -.
DR   EMBL; D50453; BAA08982.1; -.
DR   EMBL; M59939; AAA22815.1; -.
DR   EMBL; M64702; AAA22816.1; -.
DR   EMBL; D30762; BAA21034.1; -.
DR   EMBL; Z99105; CAB12142.1; -.
DR   PIR; S35517; S35517.
DR   HSSP; P14687; 1AMU.
DR   SubtiList; BG10168; srfAA.
DR   InterPro; IPR000873; AMP-bind.
DR   InterPro; IPR001242; Condensatn.
DR   InterPro; IPR003880; Ppantne_attach.
DR   Pfam; PF00501; AMP-binding; 3.
DR   Pfam; PF00550; pp-binding; 3.
DR   Pfam; PF00668; Condensation; 4.
DR   PRINTS; PR00154; AMPBINDING.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; 3.
DR   PROSITE; PS00455; AMP_BINDING; 3.
DR   PROSITE; PS50075; ACP_DOMAIN; 3.
```

```
KW   Ligase; Antibiotic biosynthesis; Phosphopantetheine; Sporulation;
KW   Multifunctional enzyme; Repeat; Complete proteome.
FT   REPEAT        ?   1047       DOMAIN 1 (GLU-ACTIVATING).
FT   REPEAT        ?   2086       DOMAIN 2 (LEU-ACTIVATING).
FT   REPEAT        ?   3114       DOMAIN 3 (D-LEU-ACTIVATING).
FT   DOMAIN      976   1043       ACYL CARRIER (ACP) 1.
FT   DOMAIN     2015   2082       ACYL CARRIER (ACP) 2.
FT   DOMAIN     3043   3109       ACYL CARRIER (ACP) 3.
FT   BINDING    1006   1006       PHOSPHOPANTETHEINE (POTENTIAL).
FT   BINDING    2045   2045       PHOSPHOPANTETHEINE (POTENTIAL).
FT   BINDING    3073   3073       PHOSPHOPANTETHEINE (POTENTIAL).
FT   CONFLICT    146    146       I -> M (IN REF. 2 AND 3).
FT   CONFLICT    151    151       L -> V (IN REF. 2 AND 3).
FT   CONFLICT    165    165       A -> G (IN REF. 2, 3 AND 4).
FT   CONFLICT    281    281       T -> Q (IN REF. 2 AND 3).
FT   CONFLICT    460    460       D -> T (IN REF. 3 AND 3).
FT   CONFLICT    540    540       P -> A (IN REF. 2 AND 3).
FT   CONFLICT    562    562       I -> Y (IN REF. 2 AND 3).
FT   CONFLICT    639    640       GS -> PT (IN REF. 2 AND 3).
FT   CONFLICT    644    644       R -> P (IN REF. 2 AND 3).
FT   CONFLICT    647    647       L -> F (IN REF. 2 AND 3).
FT   CONFLICT    649    649       A -> P (IN REF. 2 AND 3).
FT   CONFLICT   1026   1026       L -> Q (IN REF. 2 AND 3).
FT   CONFLICT   1065   1116       SLAQQRTYIVSQFEDAGVGYNMPAAAILEGPLDIQKLERAF
FT                                QGLIRRHESLR -> HWHSSEHISSASSRMRESAIHASSSN
FT                                SEGFRYSKAGARISGINPTPRVIE (IN REF. 2 AND
FT                                3).
FT   CONFLICT   1132   1133       DS -> VC (IN REF. 2 AND 3).
FT   CONFLICT   1164   1164       R -> A (IN REF. 2 AND 3).
FT   CONFLICT   1458   1458       V -> L (IN REF. 2 AND 3).
FT   CONFLICT   1850   1850       S -> T (IN REF. 2 AND 3).
FT   CONFLICT   1894   1894       D -> T (IN REF. 2 AND 3).
FT   CONFLICT   1911   1911       R -> A (IN REF. 2 AND 3).
FT   CONFLICT   1973   1980       LRCLSKWT -> PAVFIQMD (IN REF. 2 AND 3).
FT   CONFLICT   2052   2052       L -> P (IN REF. 2 AND 3).
FT   CONFLICT   2218   2218       H -> Q (IN REF. 2 AND 3).
FT   CONFLICT   2265   2266       QQ -> HR (IN REF. 2 AND 3).
FT   CONFLICT   2291   2291       V -> L (IN REF. 2 AND 3).
FT   CONFLICT   2349   2349       S -> C (IN REF. 2 AND 3).
FT   CONFLICT   2428   2428       R -> S (IN REF. 2 AND 3).
FT   CONFLICT   2611   2612       RC -> AV (IN REF. 2 AND 3).
FT   CONFLICT   2685   2685       E -> ENPE (IN REF. 2 AND 3).
FT   CONFLICT   2755   2755       T -> S (IN REF. 2 AND 3).
FT   CONFLICT   2896   2897       SP -> TA (IN REF. 2 AND 3).
FT   CONFLICT   3025   3025       P -> N (IN REF. 2 AND 3).
FT   CONFLICT   3096   3096       F -> N (IN REF. 2 AND 3).
FT   CONFLICT   3271   3271       A -> S (IN REF. 2 AND 3).
FT   CONFLICT   3316   3316       R -> S (IN REF. 2 AND 3).
FT   CONFLICT   3451   3451       Y -> S (IN REF. 2 AND 3).
FT   CONFLICT   3483   3484       DE -> HQ (IN REF. 2 AND 3).
FT   CONFLICT   3487   3490       DAGL -> HPPF (IN REF. 2 AND 3).
FT   CONFLICT   3493   3493       R -> T (IN REF. 2 AND 3).
FT   CONFLICT   3495   3495       E -> H (IN REF. 2 AND 3).
FT   CONFLICT   3499   3500       GQ -> PH (IN REF. 2 AND 3).
FT   CONFLICT   3507   3507       R -> E (IN REF. 2 AND 3).
SQ   SEQUENCE   3587 AA;  402424 MW;  A1E2DABFF93EDE3A CRC64;

  Query Match             16.6%;  Score 71;  DB 1;  Length 3587;
  Best Local Similarity   35.0%;  Pred. No. 19;
  Matches   14;  Conservative    8;  Mismatches   18;  Indels    0;  Gaps    0;

Qy         39 FRDLGFDSLTAVELRNRLSRTTGLRLPTTLAFDHPNPTTL 78
              | : |  || |: |  :: : ||: :|    |:||  | |
Db        999 FFETGGHSLKAMTLLTKIHKETGIEIPLQFLFEHPTITAL 1038


RESULT 13
ACP_SYNY3
ID   ACP_SYNY3      STANDARD;      PRT;    77 AA.
AC   P20804; P73284;
DT   01-FEB-1991 (Rel. 17, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Acyl carrier protein (ACP).
GN   ACPP OR ACP OR SSL2084.
OS   Synechocystis sp. (strain PCC 6803).
OC   Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
OX   NCBI_TaxID=1148;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=97061201; PubMed=8905231;
RA   Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,
RA   Miyajima N., Hirosawa M., Sugiura M., Sasamoto S., Kimura T.,
RA   Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Naruo K.,
RA   Okumura S., Shimpo S., Takeuchi C., Wada T., Watanabe A.,
RA   Yamada M., Yasuda M., Tabata S.;
RT   "Sequence analysis of the genome of the unicellular cyanobacterium
RT   Synechocystis sp. strain PCC6803. II. Sequence determination of the
RT   entire genome and assignment of potential protein-coding regions.";
RL   DNA Res. 3:109-136(1996).
RN   [2]
RP   SEQUENCE OF 1-61, AND SEQUENCE OF 21-45 FROM N.A.
RC   STRAIN=ATCC 27184;
RX   MEDLINE=91065389; PubMed=2123456;
RA   Froehlich J.E., Poorman R., Reardon E., Barnum S.R., Jaworski J.G.;
RT   "Purification and characterization of acyl carrier protein from two
RT   cyanobacteria species.";
RL   Eur. J. Biochem. 193:817-825(1990).
CC   -!- FUNCTION: Carrier of the growing fatty acid chain in fatty
CC       acid biosynthesis.
CC   -!- PATHWAY: De novo fatty acid biosynthesis.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC   -!- PTM: 4'-phosphopantetheine is transferred from CoA to a specific
CC       serine of apo-ACP by acpS. This modification is essential for
CC       activity because fatty acids are bound in thioester linkage to the
CC       sulfhydryl of the prosthetic group.
CC   -!- SIMILARITY: CONTAINS 1 ACYL CARRIER DOMAIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; D90905; BAA17312.1; -.
DR   PIR; S13820; S13820.
DR   HSSP; P02901; 1ACP.
DR   InterPro; IPR003231; Acyl_carrier.
DR   InterPro; IPR003880; Ppantne_attach.
DR   Pfam; PF00550; pp-binding; 1.
DR   ProDom; PD000887; Acyl_carrier; 1.
DR   TIGRFAMs; TIGR00517; acyl_carrier; 1.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; 1.
DR   PROSITE; PS50075; ACP_DOMAIN; 1.
KW   Lipid synthesis; Fatty acid biosynthesis; Phosphopantetheine;
KW   Complete proteome.
FT   BINDING      38     38       PHOSPHOPANTETHEINE (BY SIMILARITY).
FT   CONFLICT      2      2       N -> D (IN REF. 2).
FT   CONFLICT     23     23       K -> G (IN REF. 2).
SQ   SEQUENCE   77 AA;  8590 MW;  FDDED4DBFD6EBD53 CRC64;

  Query Match             16.2%;  Score 69.5;  DB 1;  Length 77;
  Best Local Similarity   41.5%;  Pred. No. 0.4;
  Matches   17;  Conservative    4;  Mismatches   19;  Indels    1;  Gaps    1;

Qy         30 PDTIPPDRAF-RDLGFDSLTAVELRNRLSRTTGLRLPTTLA 69
              || : ||  |  ||| |||  |||    |     : :|  :|
Db         21 PDKVTPDATFAEDLGADSLDTVELVMALEEEFDIEIPDEVA 61


RESULT 14
KRAF_CAEEL
ID   KRAF_CAEEL     STANDARD;      PRT;   813 AA.
AC   Q07292; Q9N4E3;
```

```
DT   01-OCT-1994 (Rel. 30, Created)
DT   15-JUN-2002 (Rel. 41, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Raf homolog serine/threonine-protein kinase (EC 2.7.1.-).
GN   LIN-45 OR RAF-1 OR Y73B6A.5.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93247635; PubMed=8483497;
RA   Han M., Golden A., Han Y., Sternberg P.W.;
RT   "C. elegans lin-45 raf gene participates in let-60 ras-stimulated
RT   vulval differentiation.";
RL   Nature 363:133-140(1993).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Bristol N2;
RA   Becker M., Graves T., Shafer S.;
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   REVISIONS.
RA   Waterston R.;
RL   Submitted (JUN-2001) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: PROTEIN KINASE THAT PARTICIPATES IN THE INDUCTION OF
CC       C.ELEGANS VULVA. ACTS DOWNSTREAM OF THE RAS PROTEIN LET-60.
CC   -!- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
CC       MIL/RAF SUBFAMILY.
CC   -!- SIMILARITY: CONTAINS 1 ZINC-DEPENDENT PHORBOL-ESTER AND DAG
CC       BINDING DOMAIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L15347; AAA28142.1; -.
DR   EMBL; AC024204; AAF36042.1; -.
DR   PIR; S33261; S33261.
DR   WormPep; Y73B6A.5; CE25585.
DR   HSSP; P04049; 1FAR.
DR   InterPro; IPR002219; DAG_PE-bind.
DR   InterPro; IPR000719; Euk_pkinase.
DR   InterPro; IPR003116; RBD.
DR   InterPro; IPR004040; STY_pkinase.
DR   InterPro; IPR002290; Ser_thr_pkinase.
DR   Pfam; PF00069; pkinase; 1.
DR   Pfam; PF00130; DAG_PE-bind; 1.
DR   Pfam; PF02196; RBD; 1.
DR   ProDom; PD000001; Euk_pkinase; 1.
DR   SMART; SM00109; C1; 1.
DR   SMART; SM00455; RBD; 1.
DR   SMART; SM00221; STYKc; 1.
DR   PROSITE; PS00479; DAG_PE_BIND_DOM_1; 1.
DR   PROSITE; PS50081; DAG_PE_BIND_DOM_2; 1.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR   PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
KW   Transferase; Serine/threonine-protein kinase; ATP-binding; Zinc;
KW   Phorbol-ester binding.
FT   DOMAIN      171    217       PHORBOL-ESTER AND DAG BINDING.
FT   DOMAIN      481    748       PROTEIN KINASE.
FT   NP_BIND     487    495       ATP (BY SIMILARITY).
FT   BINDING     507    507       ATP (BY SIMILARITY).
FT   ACT_SITE    602    602       BY SIMILARITY.
FT   CONFLICT    801    801       A -> R (IN REF. 1).
SQ   SEQUENCE   813 AA;  90406 MW;  6376E968D11A9E49 CRC64;

  Query Match             16.0%;  Score 68.5;  DB 1;  Length 813;
  Best Local Similarity   32.9%;  Pred. No. 6.9;
  Matches   26;  Conservative    8;  Mismatches   30;  Indels   15;  Gaps    5;

Qy          3 GQTHEQQHTTLLALVRSHIATVLGHTTPDTIPPDRAFRDLGFDSLTAVELRNRLSRTTGL 62
              ||:     :|  :|| :|: | |  | | : || :     ||        ||| ||: |
Db        408 GQSPNVSGSTTSSLVAAHLHT-LPLTPPQSAPPQKI--SPGF-------FRNR-SRSPGE 456

Qy         63 RLPTTLAFDHPNPTTLTHH 81
              ||        | |    ||
Db        457 RLDA----QRPRPPQKPHH 471


RESULT 15
FAS_RAT
ID   FAS_RAT        STANDARD;      PRT;  2505 AA.
AC   P12785; Q64717; O09187; O09190;
DT   01-OCT-1989 (Rel. 12, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   Fatty acid synthase (EC 2.3.1.85) [Includes: EC 2.3.1.38; EC 2.3.1.39;
DE   EC 2.3.1.41; EC 1.1.1.100; EC 4.2.1.61; EC 1.3.1.10; EC 3.1.2.14].
GN   FASN.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=89240686; PubMed=2717611;
RA   Amy C.M., Witkowski A., Naggert J., Williams B., Randhawa Z.,
RA   Smith S.;
RT   "Molecular cloning and sequencing of cDNAs encoding the entire rat
RT   fatty acid synthase.";
RL   Proc. Natl. Acad. Sci. U.S.A. 86:3114-3118(1989).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Sprague-Dawley; TISSUE=Liver;
RX   MEDLINE=93075999; PubMed=1339331;
RA   Beck K.F., Schreglmann R., Stathopulos I., Klein H., Hoch J.,
RA   Schweizer M.;
RT   "The fatty acid synthase (FAS) gene and its promoter in Rattus
RT   norvegicus.";
RL   DNA Seq. 2:359-386(1992).
RN   [3]
RP   SEQUENCE OF 75-2505 FROM N.A.
RC   STRAIN=Sprague-Dawley; TISSUE=Mammary gland;
RX   MEDLINE=89128431; PubMed=2915923;
RA   Schweizer M., Takabeyashi K., Beck K.F., Schreglmann R.;
RT   "Rat mammary gland fatty acid synthase: localization of the
RT   constituent domains and two functional polyadenylation/termination
RT   signals in the cDNA.";
RL   Nucleic Acids Res. 17:567-586(1989).
RN   [4]
RP   SEQUENCE OF 2085-2505 FROM N.A.
RC   TISSUE=Mammary gland;
RX   MEDLINE=88087240; PubMed=2891707;
RA   Naggert J., Witkowski A., Mikkelsen J., Smith S.;
RT   "Molecular cloning and sequencing of a cDNA encoding the thioesterase
RT   domain of the rat fatty acid synthetase.";
RL   J. Biol. Chem. 263:1146-1150(1988).
RN   [5]
RP   SEQUENCE OF 1921-2324 FROM N.A.
RC   TISSUE=Mammary gland;
RX   MEDLINE=87246646; PubMed=3109907;
RA   Witlowski A., Naggert J., Mikkelsen J., Smith S.;
RT   "Molecular cloning and sequencing of a cDNA encoding the acyl carrier
RT   protein and its flanking domains in the mammalian fatty acid
RT   synthetase.";
RL   Eur. J. Biochem. 165:601-606(1987).
CC   -!- FUNCTION: FATTY ACID SYNTHETASE CATALYZES THE FORMATION OF
CC       LONG-CHAIN FATTY ACIDS FROM ACETYL-COA, MALONYL-COA AND NADPH.
CC       THIS MULTIFUNCTIONAL PROTEIN HAS 7 CATALYTIC ACTIVITIES AND AN
CC       ACYL CARRIER PROTEIN.
```

```
CC   -!- CATALYTIC ACTIVITY: Acetyl-CoA + N malonyl-CoA + 2N NADPH = a
CC       long-chain fatty acid + (N+1) CoA + N CO(2) + 2N NADP(+).
CC   -!- CATALYTIC ACTIVITY: Acetyl-CoA + [acyl-carrier protein] = CoA +
CC       acetyl-[acyl-carrier protein].
CC   -!- CATALYTIC ACTIVITY: Malonyl-CoA + [acyl-carrier protein] = CoA +
CC       malonyl-[acyl-carrier protein].
CC   -!- CATALYTIC ACTIVITY: Acyl-[acyl-carrier protein] + malonyl-[acyl-
CC       carrier protein] = 3-oxoacyl-[acyl-carrier protein] + CO(2) +
CC       [acyl-carrier protein].
CC   -!- CATALYTIC ACTIVITY: (3R)-3-hydroxyacyl-[acyl-carrier protein] +
CC       NADP(+) = 3-oxoacyl-[acyl-carrier protein] + NADPH.
CC   -!- CATALYTIC ACTIVITY: (3R)-3-hydroxypalmitoyl-[acyl-carrier protein]
CC       = 2-hexadecenoyl-[acyl-carrier protein] + H(2)O.
CC   -!- CATALYTIC ACTIVITY: Acyl-[acyl-carrier protein] + NADP(+) = trans-
CC       2,3-dehydroacyl-[acyl-carrier protein] + NADPH.
CC   -!- CATALYTIC ACTIVITY: Oleoyl-[acyl-carrier protein] + H(2)O = [acyl-
CC       carrier protein] + oleate.
CC   -!- SUBUNIT: HOMODIMER, WHICH IS ARRANGED IN A HEAD TO TAIL FASHION.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; M76767; AAA57219.1; -.
DR   EMBL; X62888; CAA44679.1; -.
DR   EMBL; X62889; CAA44680.1; -.
DR   EMBL; X13415; CAA31780.1; -.
DR   EMBL; X13527; CAA31882.1; -.
DR   EMBL; J03514; AAA41144.1; -.
DR   PIR; A30313; XYRTFA.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR002085; Adh_zn_family.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR003880; Ppantne_attach.
DR   InterPro; IPR000051; SAM_bind.
DR   InterPro; IPR001031; Thioesterase.
DR   Pfam; PF00107; adh_zinc; 1.
DR   Pfam; PF00109; ketoacyl-synt; 1.
DR   Pfam; PF00550; pp-binding; 1.
DR   Pfam; PF00698; Acyl_transf; 1.
DR   Pfam; PF00975; Thioesterase; 1.
DR   Pfam; PF02801; ketoacyl-synt_C; 1.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; 1.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 1.
DR   PROSITE; PS50075; ACP_DOMAIN; 1.
KW   Fatty acid biosynthesis; Multifunctional enzyme; Phosphopantetheine;
KW   Hydrolase; Oxidoreductase; Transferase; Lyase; NADP;
KW   Pyridoxal phosphate.
FT   DOMAIN        1    413       BETA-KETOACYL SYNTHASE.
FT   DOMAIN      429    817       ACYL AND MALONYL TRANSFERASES.
FT   DOMAIN     1629   1857       ENOYL REDUCTASE.
FT   DOMAIN     1858   2113       BETA-KETOACYL REDUCTASE.
FT   DOMAIN     2118   2174       ACYL CARRIER (ACP).
FT   DOMAIN     2202   2505       THIOESTERASE.
FT   ACT_SITE    161    161       BETA-KETOACYL SYNTHASE (BY SIMILARITY).
FT   ACT_SITE    581    581       MALONYLTRANSFERASE (BY SIMILARITY).
FT   NP_BIND    1665   1682       NADP (ER).
FT   BINDING    1698   1698       PYRIDOXAL PHOSPHATE (BY SIMILARITY).
FT   NP_BIND    1765   1780       NADP (KR).
FT   BINDING    2151   2151       PHOSPHOPANTETHEINE (BY SIMILARITY).
FT   ACT_SITE   2302   2302       THIOESTERASE (BY SIMILARITY).
FT   ACT_SITE   2475   2475       THIOESTERASE (BY SIMILARITY).
FT   ACT_SITE    878    878       BETA-HYDROXYACYL DEHYDRATASE (BY
FT                                SIMILARITY).
FT   CONFLICT    871    871       S -> P (IN REF. 3).
FT   CONFLICT   1967   1968       MV -> IL (IN REF. 5).
FT   CONFLICT   2085   2085       C -> P (IN REF. 4).
FT   CONFLICT   2106   2106       A -> V (IN REF. 1 AND 5).
FT   CONFLICT   2296   2296       Y -> H (IN REF. 1 AND 5).
```

```
SQ   SEQUENCE   2505 AA;  272647 MW;  5810EC13D37F3114 CRC64;

  Query Match           16.0%;  Score 68.5;  DB 1;  Length 2505;
  Best Local Similarity 42.0%;  Pred. No. 24;
  Matches   21;  Conservative    6;  Mismatches   22;  Indels    1;  Gaps    1;

Qy         16 LVRSHIATVLGHTTPDTIPPDRAFRDLGFDSLTAVELRNRLSRTTGLRLP 65
              ||:: :| :||      |  | :  ||| |||  ||:|  | |   | ||
Db       2122 LVKA-VAHILGIRDLAGINLDSSLADLGLDSLMGVEVRQILEREHDLVLP 2170


Search completed: June 17, 2003, 13:02:41
Job time : 1.89424 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 17, 2003, 12:53:02 ; Search time 3.78986 Seconds
(without alignments)
4458.180 Million cell updates/sec

Title: US-09-914-286-3_COPY_3805_3886
Perfect score: 429
Sequence: 1 LAGQTHEQQHTTLLALVRSH.....RLPTTLAFDHPNPTTLTHHL 82

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_21:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	429	100.0	3972	2	Q9S0R8	Q9s0r8 streptomyce
2	429	100.0	5532	2	Q9S0R4	Q9s0r4 streptomyce
3	270	62.9	3729	2	O33956	O33956 streptomyce
4	258	60.1	3613	2	Q93HJ1	Q93hj1 streptomyce
5	254	59.2	4881	2	Q9S0R3	Q9s0r3 streptomyce
6	253	59.0	3739	2	Q9ZGI4	Q9zgi4 streptomyce
7	249	58.0	4151	16	O53490	O53490 mycobacteri
8	246	57.3	1762	2	O52546	O52546 amycolatops
9	246	57.3	1763	2	O52790	O52790 amycolatops
10	246	57.3	3524	2	Q93H86	Q93h86 streptomyce
11	246	57.3	6048	2	Q93H87	Q93h87 streptomyce
12	245	57.1	5644	2	Q93NX8	Q93nx8 streptomyce
13	245	57.1	6239	2	Q9S0R7	Q9s0r7 streptomyce
14	244	56.9	4685	2	Q93HJ2	Q93hj2 streptomyce
15	243	56.6	3626	2	Q9F7T9	Q9f7t9 streptomyce
16	243	56.6	10917	2	Q93NW6	Q93nw6 streptomyce
17	241	56.2	6797	2	Q9X993	Q9x993 streptomyce
18	240	55.9	3816	2	Q9KIV3	Q9kiv3 streptomyce
19	239	55.7	6146	2	Q93HJ5	Q93hj5 streptomyce
20	237	55.2	1728	2	O54591	O54591 amycolatops
21	236	55.0	902	2	Q93AL5	Q93al5 streptomyce
22	236	55.0	3654	2	O30766	O30766 streptomyce
23	236	55.0	4150	2	Q9KIV4	Q9kiv4 streptomyce
24	236	55.0	11096	2	Q9L4W3	Q9l4w3 streptomyce
25	235	54.8	9477	2	Q9L4X3	Q9l4x3 streptomyce
26	234	54.5	1562	2	Q9ZGI3	Q9zgi3 streptomyce
27	234	54.5	2100	2	O30481	O30481 streptomyce
28	234	54.5	6145	2	Q93H84	Q93h84 streptomyce
29	233	54.3	3562	2	Q9F829	Q9f829 micromonosp
30	233	54.3	5060	2	O52545	O52545 amycolatops
31	233	54.3	5069	2	O52789	O52789 amycolatops
32	231	53.8	3352	2	Q93H83	Q93h83 streptomyce
33	231	53.8	3576	16	Q9EX53	Q9ex53 streptomyce
34	229	53.4	5435	2	Q9L4X2	Q9l4x2 streptomyce
35	227	52.9	9507	2	Q9EWA1	Q9ewa1 streptomyce
36	226	52.7	1762	2	O30480	O30480 streptomyce
37	226	52.7	3170	2	Q9ALM4	Q9alm4 saccharopol
38	225	52.4	3939	2	Q93HJ3	Q93hj3 streptomyce
39	225	52.4	4498	2	Q93HE5	Q93he5 streptomyce
40	224	52.2	3201	2	Q9F828	Q9f828 micromonosp
41	224	52.2	3413	2	O54593	O54593 amycolatops
42	223	52.0	1835	2	Q93H85	Q93h85 streptomyce
43	222	51.7	1839	2	O30765	O30765 streptomyce
44	222	51.7	3192	2	Q9L4W4	Q9l4w4 streptomyce
45	222	51.7	4290	2	Q9WXC0	Q9wxc0 micromonosp

ALIGNMENTS

RESULT 1
Q9S0R8
ID Q9S0R8 PRELIMINARY; PRT; 3972 AA.
AC Q9S0R8;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Type I polyketide synthase AVES 1.
GN AVEA1.
OS Streptomyces avermitilis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=33903;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99380548; PubMed=10449723;
RA Ikeda H., Nonomiya T., Usami M., Ohta T., Omura S.;
RT "Organization of the biosynthetic gene cluster for the polyketide
RT anthelmintic macrolide avermectin in Streptomyces avermitilis.";
RL Proc. Natl. Acad. Sci. U.S.A. 96:9509-9514(1999).
DR EMBL; AB032367; BAA84474.1; -.
DR InterPro; IPR001227; Ac_transferase.
DR InterPro; IPR000794; Ketoacyl-synt.
DR InterPro; IPR003880; Ppantne_attach.
DR InterPro; IPR000834; Zn_carbopept.
DR Pfam; PF00698; Acyl_transf; 3.
DR Pfam; PF00109; ketoacyl-synt; 2.
DR Pfam; PF02801; ketoacyl-synt_C; 2.
DR Pfam; PF00550; pp-binding; 3.
DR PROSITE; PS50075; ACP_DOMAIN; 3.
DR PROSITE; PS00606; B_KETOACYL_SYNTHASE; 2.
DR PROSITE; PS00133; CARBOXYPEPT_ZN_2; UNKNOWN_2.
DR PROSITE; PS00012; PHOSPHOPANTETHEINE; 2.
KW Phosphopantetheine; Transferase.
SQ SEQUENCE 3972 AA; 416852 MW; 2A293695B032B1C3 CRC64;

Query Match 100.0%; Score 429; DB 2; Length 3972;
Best Local Similarity 100.0%; Pred. No. 1e-38;
Matches 82; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 LAGQTHEQQHTTLLALVRSHIATVLGHTTPDTIPPDRAFRDLGFDSLTAVELRNRLSRTT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3805 LAGQTHEQQHTTLLALVRSHIATVLGHTTPDTIPPDRAFRDLGFDSLTAVELRNRLSRTT 3864

Qy         61 GLRLPTTLAFDHPNPTTLTHHL 82
              |||||||||||||||||||||||
Db       3865 GLRLPTTLAFDHPNPTTLTHHL 3886


RESULT 2
Q9S0R4
ID   Q9S0R4      PRELIMINARY;      PRT;  5532 AA.
AC   Q9S0R4;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Type I polyketide synthase AVES 3.
GN   AVEA3.
OS   Streptomyces avermitilis.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=33903;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99380548; PubMed=10449723;
RA   Ikeda H., Nonomiya T., Usami M., Ohta T., Omura S.;
RT   "Organization of the biosynthetic gene cluster for the polyketide
RT   anthelmintic macrolide avermectin in Streptomyces avermitilis.";
RL   Proc. Natl. Acad. Sci. U.S.A. 96:9509-9514(1999).
DR   EMBL; AB032367; BAA84478.1; -.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR003880; Ppantne_attach.
DR   InterPro; IPR000834; Zn_carbOpept.
DR   Pfam; PF00698; Acyl_transf; 3.
DR   Pfam; PF00109; ketoacyl-synt; 3.
DR   Pfam; PF02801; ketoacyl-synt_C; 3.
DR   Pfam; PF00550; pp-binding; 3.
DR   PROSITE; PS50075; ACP_DOMAIN; 3.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 3.
DR   PROSITE; PS00133; CARBOXYPEPT_ZN_2; UNKNOWN_1.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; 3.
KW   Phosphopantetheine; Transferase.
SQ   SEQUENCE   5532 AA;  575193 MW;  EC7B2069D44A4057 CRC64;

  Query Match           100.0%;  Score 429;  DB 2;  Length 5532;
  Best Local Similarity  100.0%;  Pred. No. 1.5e-38;
  Matches   82;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 LAGQTHEQQHTTLLALVRSHIATVLGHTTPDTIPPDRAFRDLGFDSLTAVELRNRLSRTT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1756 LAGQTHEQQHTTLLALVRSHIATVLGHTTPDTIPPDRAFRDLGFDSLTAVELRNRLSRTT 1815

Qy         61 GLRLPTTLAFDHPNPTTLTHHL 82
              |||||||||||||||||||||||
Db       1816 GLRLPTTLAFDHPNPTTLTHHL 1837


RESULT 3
O33956
ID   O33956      PRELIMINARY;      PRT;  3729 AA.
AC   O33956;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Tylactone synthase modules 4 & 5.
GN   TYLG.
OS   Streptomyces fradiae.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=1906;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   DeHoff B.S., Sutton K.L., Rosteck P.R. Jr.;
RT   "Sequence of Streptomyces fradiae tylactone synthase gene tylG.";
RL   Submitted (NOV-1996) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; U78289; AAB66506.1; -.
DR   InterPro; IPR002106; AAtRNA_ligaseII.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR002085; Adh_zn_family.
DR   InterPro; IPR001899; Gram_pos_anchor.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR003880; Ppantne_attach.
DR   InterPro; IPR002364; QOR_zeta_crystal.
DR   InterPro; IPR001680; WD40.
DR   Pfam; PF00698; Acyl_transf; 2.
DR   Pfam; PF00107; adh_zinc; 1.
DR   Pfam; PF00109; ketoacyl-synt; 2.
DR   Pfam; PF02801; ketoacyl-synt_C; 2.
DR   Pfam; PF00550; pp-binding; 2.
DR   PROSITE; PS00339; AA_TRNA_LIGASE_II_2; UNKNOWN_1.
DR   PROSITE; PS50075; ACP_DOMAIN; 2.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 2.
DR   PROSITE; PS00343; GRAM_POS_ANCHORING; UNKNOWN_2.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; UNKNOWN_2.
DR   PROSITE; PS01162; QOR_ZETA_CRYSTAL; 1.
DR   PROSITE; PS00678; WD_REPEATS_1; UNKNOWN_1.
KW   Phosphopantetheine; Transferase.
SQ   SEQUENCE   3729 AA;  389375 MW;  19483A58E027BAC1 CRC64;

  Query Match           62.9%;  Score 270;  DB 2;  Length 3729;
  Best Local Similarity  63.4%;  Pred. No. 5.9e-21;
  Matches   52;  Conservative    9;  Mismatches   21;  Indels    0;  Gaps    0;

Qy          1 LAGQTHEQQHTTLLALVRSHIATVLGHTTPDTIPPDRAFRDLGFDSLTAVELRNRLSRTT 60
              | | | :: |  :|| :|:| | |||| : |:|| ||||:||||||||||:|||||  |
Db       3547 LTGLTAQEGHALVLAEIRAHAAAVLGHGSDDSIPEDRAFKDLGFDSLTAVEMRNRLSAAT 3606

Qy         61 GLRLPTTLAFDHPNPTTLTHHL 82
              ||||| || |||| |  |  ||
Db       3607 GLRLPATLVFDHPTPGELAGHL 3628


RESULT 4
Q93HJ1
ID   Q93HJ1      PRELIMINARY;      PRT;  3613 AA.
AC   Q93HJ1;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Modular polyketide synthase.
GN   OLMA7.
OS   Streptomyces avermitilis.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=33903;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21477403; PubMed=11572948;
RA   Omura S., Ikeda H., Ishikawa J., Hanamoto A., Takahashi C.,
RA   Shinose M., Takahashi Y., Horikawa H., Nakazawa H., Osonoe T.,
RA   Kikuchi H., Shiba T., Sakaki Y., Hattori M.;
RT   "Genome sequence of an industrial microorganism Streptomyces
RT   avermitilis: Deducing the ability of producing secondary
RT   metabolites.";
RL   Proc. Natl. Acad. Sci. U.S.A. 98:12215-12220(2001).
DR   EMBL; AB070940; BAB69196.1; -.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR002198; ADH_short.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR003880; Ppantne_attach.
DR   InterPro; IPR001031; Thioesterase.
DR   Pfam; PF00698; Acyl_transf; 2.
DR   Pfam; PF00106; adh_short; 1.
DR   Pfam; PF00109; ketoacyl-synt; 2.
```

```
DR   Pfam; PF02801; ketoacyl-synt_C; 2.
DR   Pfam; PF00550; pp-binding; 2.
DR   Pfam; PF00975; Thioesterase; 1.
DR   PROSITE; PS50075; ACP_DOMAIN; 2.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; UNKNOWN_2.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; UNKNOWN_2.
KW   Phosphopantetheine.
SQ   SEQUENCE   3613 AA;  379236 MW;  6FAAE5CAB158E442 CRC64;

  Query Match             60.1%;  Score 258;  DB 2;  Length 3613;
  Best Local Similarity   65.9%;  Pred. No. 1.3e-19;
  Matches   54;  Conservative    3;  Mismatches   25;  Indels    0;  Gaps    0;

Qy          1 LAGQTHEQQHTTLLALVRSHIATVLGHTTPDTIPPDRAFRDLGFDSLTAVELRNRLSRTT 60
              ||| |  :||  || ||||  |||| | | | |    | ||||||||||:||||||:  |
Db       1462 LAGLTAPEQHDRLLELVRSEAATVLTHRTTDDITAGRPFRDLGFDSLTAMELRNRLNTAT 1521

Qy         61 GLRLPTTLAFDHPNPTTLTHHL 82
              ||||| || |||| |  |  ||
Db       1522 GLRLPATLVFDHPTPQRLAGHL 1543


RESULT 5
Q9S0R3
ID   Q9S0R3      PRELIMINARY;      PRT;  4881 AA.
AC   Q9S0R3;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Type I polyketide synthase AVES 4.
GN   AVEA4.
OS   Streptomyces avermitilis.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=33903;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99380548; PubMed=10449723;
RA   Ikeda H., Nonomiya T., Usami M., Ohta T., Omura S.;
RT   "Organization of the biosynthetic gene cluster for the polyketide
RT   anthelmintic macrolide avermectin in Streptomyces avermitilis.";
RL   Proc. Natl. Acad. Sci. U.S.A. 96:9509-9514(1999).
DR   EMBL; AB032367; BAA84479.1; -.
DR   HSSP; P25715; 1MLA.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR003880; Ppantne_attach.
DR   InterPro; IPR001031; Thioesterase.
DR   Pfam; PF00698; Acyl_transf; 3.
DR   Pfam; PF00109; ketoacyl-synt; 3.
DR   Pfam; PF02801; ketoacyl-synt_C; 3.
DR   Pfam; PF00550; pp-binding; 3.
DR   Pfam; PF00975; Thioesterase; 1.
DR   PROSITE; PS50075; ACP_DOMAIN; 3.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 3.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; 2.
KW   Phosphopantetheine; Transferase.
SQ   SEQUENCE   4881 AA;  510303 MW;  F6568C1F01ACD37A CRC64;

  Query Match             59.2%;  Score 254;  DB 2;  Length 4881;
  Best Local Similarity   62.2%;  Pred. No. 5e-19;
  Matches   51;  Conservative    6;  Mismatches   25;  Indels    0;  Gaps    0;

Qy          1 LAGQTHEQQHTTLLALVRSHIATVLGHTTPDTIPPDRAFRDLGFDSLTAVELRNRLSRTT 60
              |||      |  |:| || :| | |||| | | : |:||||||||||||||||||:  |
Db       2525 LAGLDPAAQAETVLDLVLTHAAAVLGHGTADAVVPERAFRDLGFDSLTAVELRNRLNTAT 2584

Qy         61 GLRLPTTLAFDHPNPTTLTHHL 82
              ||| | || |||| |  |  |:
Db       2585 GLRFPRTLVFDHPRPVALAAHI 2606


RESULT 6
Q9ZGI4
ID   Q9ZGI4      PRELIMINARY;      PRT;  3739 AA.
AC   Q9ZGI4;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Type I polyketide synthase PikAII.
GN   PIKAII.
OS   Streptomyces venezuelae.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=54571;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC15439;
RX   MEDLINE=98445333; PubMed=9770448;
RA   Xue Y., Zhao L., Liu H.w., Sherman D.H.;
RT   "A gene cluster for macrolide antibiotic biosynthesis in streptomyces
RT   venezuelae: architecture of metabolic diversity.";
RL   Proc. Natl. Acad. Sci. U.S.A. 95:12111-12116(1998).
DR   EMBL; AF079138; AAC69330.1; -.
DR   HSSP; Q02054; 2AF8.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR002085; Adh_zn_family.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR003880; Ppantne_attach.
DR   InterPro; IPR002364; QOR_zeta_crystal.
DR   InterPro; IPR000169; SHprot_acsite.
DR   Pfam; PF00698; Acyl_transf; 2.
DR   Pfam; PF00107; adh_zinc; 1.
DR   Pfam; PF00109; ketoacyl-synt; 2.
DR   Pfam; PF02801; ketoacyl-synt_C; 2.
DR   Pfam; PF00550; pp-binding; 2.
DR   PROSITE; PS50075; ACP_DOMAIN; 2.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 2.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; 2.
DR   PROSITE; PS01162; QOR_ZETA_CRYSTAL; 1.
DR   PROSITE; PS00639; THIOL_PROTEASE_HIS; UNKNOWN_1.
KW   Phosphopantetheine; Transferase.
SQ   SEQUENCE   3739 AA;  387170 MW;  3D3910824DA5B080 CRC64;

  Query Match             59.0%;  Score 253;  DB 2;  Length 3739;
  Best Local Similarity   62.2%;  Pred. No. 4.8e-19;
  Matches   51;  Conservative    7;  Mismatches   24;  Indels    0;  Gaps    0;

Qy          1 LAGQTHEQQHTTLLALVRSHIATVLGHTTPDTIPPDRAFRDLGFDSLTAVELRNRLSRTT 60
              ||  | : :   |  |||:|:|||||| ||  :  :|||| ||||||||||||||:  |
Db       3560 LAAMTPDDRVAHLRDLVRTHVATVLGHGTPSRVDLERAFRDTGFDSLTAVELRNRLNAAT 3619

Qy         61 GLRLPTTLAFDHPNPTTLTHHL 82
              ||||| || |||| |  |  ||
Db       3620 GLRLPATLVFDHPTPGELAGHL 3641


RESULT 7
O53490
ID   O53490      PRELIMINARY;      PRT;  4151 AA.
AC   O53490;
DT   01-JUN-1998 (TrEMBLrel. 06, Created)
DT   01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Polyketide synthase.
GN   PKS12 OR RV2048C OR MT2108 OR MTV018.35C.
OS   Mycobacterium tuberculosis.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1773;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H37RV;
RX   MEDLINE=98295987; PubMed=9634230;
RA   Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA   Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekaia F.,
```

```
RA   Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA   Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA   Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA   Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA   Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA   Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT   "Deciphering the biology of Mycobacterium tuberculosis from the
RT   complete genome sequence.";
RL   Nature 393:537-544(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CDC 1551 / OSHKOSH;
RA   Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA   Peterson J., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey E.,
RA   Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,
RA   Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA   Bishai W.;
RT   "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT   laboratory strains.";
RL   Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AL021899; CAA17262.1; -.
DR   EMBL; AE007061; AAK46387.1; -.
DR   HSSP; P25715; 1MLA.
DR   TIGR; MT2108; -.
DR   TubercuList; Rv2048c; -.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR002085; Adh_zn_family.
DR   InterPro; IPR001899; Gram_pos_anchor.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR003880; Ppantne_attach.
DR   Pfam; PF00698; Acyl_transf; 2.
DR   Pfam; PF00107; adh_zinc; 2.
DR   Pfam; PF00109; ketoacyl-synt; 2.
DR   Pfam; PF02801; ketoacyl-synt_C; 2.
DR   Pfam; PF00550; pp-binding; 2.
DR   PROSITE; PS50075; ACP_DOMAIN; 2.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 2.
DR   PROSITE; PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; UNKNOWN_2.
KW   Phosphopantetheine; Transferase; Complete proteome.
FT   CONFLICT   2147   2147       H -> Q (IN REF. 2).
FT   CONFLICT   2260   2261       DI -> TV (IN REF. 2).
FT   CONFLICT   2268   2268       W -> H (IN REF. 2).
FT   CONFLICT   2272   2272       S -> A (IN REF. 2).
FT   CONFLICT   2279   2280       AF -> PY (IN REF. 2).
FT   CONFLICT   2282   2283       AA -> GR (IN REF. 2).
FT   CONFLICT   2287   2287       T -> V (IN REF. 2).
FT   CONFLICT   2289   2289       F -> W (IN REF. 2).
FT   CONFLICT   3004   3004       S -> L (IN REF. 2).
FT   CONFLICT   3649   3649       P -> A (IN REF. 2).
SQ   SEQUENCE   4151 AA;  431577 MW;  536F644F111D7A5D CRC64;

  Query Match             58.0%;  Score 249;  DB 16;  Length 4151;
  Best Local Similarity   59.8%;  Pred. No. 1.5e-18;
  Matches   49;  Conservative    9;  Mismatches   24;  Indels    0;  Gaps    0;

Qy          1 LAGQTHEQQHTTLLALVRSHIATVLGHTTPDTIPPDRAFRDLGFDSLTAVELRNRLSRTT 60
              | |    :||  || ||| |||||||: ||: | ||:||:|||||||||||:||||   |
Db       1953 LHGLPEAEQHAVLLGLVRLHIATVLGNITPEAIDPDKAFQDLGFDSLTAVEMRNRLKSAT 2012

Qy         61 GLRLPTTLAFDHPNPTTLTHHL 82
              || |  || ||:| |  |  ::
Db       2013 GLSLSPTLIFDYPTPNRLASYI 2034


RESULT 8
O52546
ID   O52546      PRELIMINARY;      PRT;  1762 AA.
AC   O52546;
DT   01-JUN-1998 (TrEMBLrel. 06, Created)
DT   01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Polyketide synthase.
GN   RIFC.
OS   Amycolatopsis mediterranei (Nocardia mediterranei).
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Pseudonocardineae; Pseudonocardiaceae; Amycolatopsis.
OX   NCBI_TaxID=33910;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S699;
RX   MEDLINE=98174059; PubMed=9512878;
RA   August P.R., Tang L., Yoon Y.J., Ning S., Mueller R., Yu T.W.,
RA   Taylor M., Hoffmann D., Kim C.G., Zhang X., Hutchinson C.R.,
RA   Floss H.G.;
RT   "Biosynthesis of the ansamycin antibiotic rifamycin: deductions from
RT   the molecular analysis of the rif biosynthetic gene cluster of
RT   Amycolatopsis mediterranei S699.";
RL   Chem. Biol. 5:69-79(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S699;
RX   MEDLINE=98165773; PubMed=9497318;
RA   Kim C.G., Yu T.W., Fryhle C.B., Handa S., Floss H.G.;
RT   "3-Amino-5-hydroxybenzoic acid synthase, the terminal enzyme in the
RT   formation of the precursor of mC7N units in rifamycin and related
RT   antibiotics.";
RL   J. Biol. Chem. 273:6030-6040(1998).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S699;
RX   MEDLINE=21201076; PubMed=11278540;
RA   Yu T.-W., Mueller R., Mueller M., Zhang X., Draeger G., Kim C.-G.,
RA   Leistner E., Floss H.G.;
RT   "Mutational analysis and reconstituted expression of the biosynthetic
RT   genes involved in the formation of 3-amino-5-hydroxybenzoic acid, the
RT   starter unit of rifamycin biosynthesis in Amycolatopsis mediterranei
RT   S699.";
RL   J. Biol. Chem. 276:12546-12555(2001).
RN   [4]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S699;
RA   August P.R., Tang L., Yoon Y.J., Ning S., Mueller R., Hutchinson C.R.,
RA   Floss H.G.;
RL   Submitted (JAN-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF040570; AAC01712.1; -.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR003880; Ppantne_attach.
DR   Pfam; PF00698; Acyl_transf; 1.
DR   Pfam; PF00109; ketoacyl-synt; 1.
DR   Pfam; PF02801; ketoacyl-synt_C; 1.
DR   Pfam; PF00550; pp-binding; 1.
DR   PROSITE; PS50075; ACP_DOMAIN; 1.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; UNKNOWN_1.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; UNKNOWN_1.
KW   Phosphopantetheine.
SQ   SEQUENCE   1762 AA;  182723 MW;  A91B5817A4F6889E CRC64;

  Query Match             57.3%;  Score 246;  DB 2;  Length 1762;
  Best Local Similarity   59.8%;  Pred. No. 1.2e-18;
  Matches   49;  Conservative    6;  Mismatches   27;  Indels    0;  Gaps    0;

Qy          1 LAGQTHEQQHTTLLALVRSHIATVLGHTTPDTIPPDRAFRDLGFDSLTAVELRNRLSRTT 60
              ||| |  :|    || |||  :| ||||      :  | ||:| |||||||:|||||||  ||
Db       1604 LAGLTVAEQEALLLDLVRGQVAVVLGHADSSGVRADAAFKDAGFDSLTSVELRNRLRETT 1663

Qy         61 GLRLPTTLAFDHPNPTTLTHHL 82
              ||:|| || ||||||  |  ||
Db       1664 GLKLPATLVFDHPNPLALARHL 1685


RESULT 9
O52790
ID   O52790      PRELIMINARY;      PRT;  1763 AA.
AC   O52790;
```

```
DT   01-JUN-1998 (TrEMBLrel. 06, Created)
DT   01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Rifamycin polyketide synthase, type 1.
OS   Amycolatopsis mediterranei (Nocardia mediterranei).
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Pseudonocardineae; Pseudonocardiaceae; Amycolatopsis.
OX   NCBI_TaxID=33910;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=LBG A3136;
RA   Schupp T., Toupet C., Engel N., Goff S.;
RT   "Cloning and sequence analysis of the putative rifamycin polyketide
RT   synthase gene cluster from Amycolatopsis mediterranei.";
RL   Submitted (DEC-1997) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AJ223012; CAA11037.1; -.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR003880; Ppantne_attach.
DR   Pfam; PF00698; Acyl_transf; 1.
DR   Pfam; PF00109; ketoacyl-synt; 1.
DR   Pfam; PF02801; ketoacyl-synt_C; 1.
DR   Pfam; PF00550; pp-binding; 1.
DR   PROSITE; PS50075; ACP_DOMAIN; 1.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; UNKNOWN_1.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; UNKNOWN_1.
KW   Phosphopantetheine.
SQ   SEQUENCE   1763 AA;  182822 MW;  DCD7FD776F006692 CRC64;

  Query Match             57.3%;  Score 246;  DB 2;  Length 1763;
  Best Local Similarity   59.8%;  Pred. No. 1.2e-18;
  Matches   49;  Conservative    6;  Mismatches   27;  Indels    0;  Gaps    0;

Qy          1 LAGQTHEQQHTTLLALVRSHIATVLGHTTPDTIPPDRAFRDLGFDSLTAVELRNRLSRTT 60
              ||| |  :|   || |||  :| ||||     :  | ||:| ||||||:|||||||  ||
Db       1605 LAGLTVAEQEALLLDLVRGQVAVVLGHADSSGVRADAAFKDAGFDSLTSVELRNRLRETT 1664

Qy         61 GLRLPTTLAFDHPNPTTLTHHL 82
              ||:|| || ||||||  |  ||
Db       1665 GLKLPATLVFDHPNPLALARHL 1686


RESULT 10
Q93H86
ID   Q93H86      PRELIMINARY;      PRT;  3524 AA.
AC   Q93H86;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Modular polyketide synthase.
GN   PTEA2.
OS   Streptomyces avermitilis.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=33903;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21477403; PubMed=11572948;
RA   Omura S., Ikeda H., Ishikawa J., Hanamoto A., Takahashi C.,
RA   Shinose M., Takahashi Y., Horikawa H., Nakazawa H., Osonoe T.,
RA   Kikuchi H., Shiba T., Sakaki Y., Hattori M.;
RT   "Genome sequence of an industrial microorganism Streptomyces
RT   avermitilis: Deducing the ability of producing secondary
RT   metabolites.";
RL   Proc. Natl. Acad. Sci. U.S.A. 98:12215-12220(2001).
DR   EMBL; AB070949; BAB69304.1; -.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR001993; Mitoch_carrier.
DR   InterPro; IPR003880; Ppantne_attach.
DR   Pfam; PF00698; Acyl_transf; 2.
DR   Pfam; PF00109; ketoacyl-synt; 2.
DR   Pfam; PF02801; ketoacyl-synt_C; 2.
DR   Pfam; PF00550; pp-binding; 2.
DR   PROSITE; PS50075; ACP_DOMAIN; 2.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; UNKNOWN_2.
DR   PROSITE; PS00215; MITOCH_CARRIER; UNKNOWN_1.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; UNKNOWN_2.
KW   Phosphopantetheine.
SQ   SEQUENCE   3524 AA;  366619 MW;  64A65759C461EBC4 CRC64;

  Query Match             57.3%;  Score 246;  DB 2;  Length 3524;
  Best Local Similarity   56.1%;  Pred. No. 2.7e-18;
  Matches   46;  Conservative   13;  Mismatches   23;  Indels    0;  Gaps    0;

Qy          1 LAGQTHEQQHTTLLALVRSHIATVLGHTTPDTIPPDRAFRDLGFDSLTAVELRNRLSRTT 60
              :||    ::   :| |||| :| ||||::  |: ||:|| ||||||||||||||||::: |
Db       3352 IAGLPESEREEAVLDLVRSEVAAVLGHSSAKTVQPDQAFGDLGFDSLTAVELRNRMNQAT 3411

Qy         61 GLRLPTTLAFDHPNPTTLTHHL 82
              ||||| || |||| |  |   :
Db       3412 GLRLPATLVFDHPTPAVLAQRV 3433


RESULT 11
Q93H87
ID   Q93H87      PRELIMINARY;      PRT;  6048 AA.
AC   Q93H87;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Modular polyketide synthase.
GN   PTEA1.
OS   Streptomyces avermitilis.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=33903;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21477403; PubMed=11572948;
RA   Omura S., Ikeda H., Ishikawa J., Hanamoto A., Takahashi C.,
RA   Shinose M., Takahashi Y., Horikawa H., Nakazawa H., Osonoe T.,
RA   Kikuchi H., Shiba T., Sakaki Y., Hattori M.;
RT   "Genome sequence of an industrial microorganism Streptomyces
RT   avermitilis: Deducing the ability of producing secondary
RT   metabolites.";
RL   Proc. Natl. Acad. Sci. U.S.A. 98:12215-12220(2001).
DR   EMBL; AB070949; BAB69303.1; -.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR004410; FabD.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR003880; Ppantne_attach.
DR   Pfam; PF00698; Acyl_transf; 4.
DR   Pfam; PF00109; ketoacyl-synt; 4.
DR   Pfam; PF02801; ketoacyl-synt_C; 4.
DR   Pfam; PF00550; pp-binding; 4.
DR   TIGRFAMs; TIGR00128; fabD; 4.
DR   PROSITE; PS50075; ACP_DOMAIN; 4.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; UNKNOWN_3.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; UNKNOWN_4.
KW   Phosphopantetheine.
SQ   SEQUENCE   6048 AA;  629420 MW;  F8673C2ED2694705 CRC64;

  Query Match             57.3%;  Score 246;  DB 2;  Length 6048;
  Best Local Similarity   62.2%;  Pred. No. 5e-18;
  Matches   51;  Conservative    4;  Mismatches   27;  Indels    0;  Gaps    0;

Qy          1 LAGQTHEQQHTTLLALVRSHIATVLGHTTPDTIPPDRAFRDLGFDSLTAVELRNRLSRTT 60
              | | |  :|   || |||   |||||||       || ||: ||||||||||:|:||:  |
Db       2446 LLGATPGEQGRILLDLVREQAATVLGHTGKGAFEADRTFRETGFDSLTAVELRHRLNTAT 2505

Qy         61 GLRLPTTLAFDHPNPTTLTHHL 82
              |||||||| |||| || |  ||
Db       2506 GLRLPTTLVFDHPTPTALARHL 2527
```

```
RESULT 12
Q93NX8
ID   Q93NX8      PRELIMINARY;      PRT;  5644 AA.
AC   Q93NX8;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   AmphJ.
GN   AMPHJ.
OS   Streptomyces nodosus.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=40318;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Caffrey P., Lynch S.V., Flood E.M., Finnan S.M., Oliynyk M.;
RT   "The amphotericin biosynthetic gene cluster from Streptomyces
RT   nodosus.";
RL   Submitted (MAR-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF357202; AAK73502.1; -.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR002085; Adh_zn_family.
DR   InterPro; IPR004410; FabD.
DR   InterPro; IPR001899; Gram_pos_anchor.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR003880; Ppantne_attach.
DR   Pfam; PF00698; Acyl_transf; 3.
DR   Pfam; PF00107; adh_zinc; 1.
DR   Pfam; PF00109; ketoacyl-synt; 3.
DR   Pfam; PF02801; ketoacyl-synt_C; 3.
DR   Pfam; PF00550; pp-binding; 3.
DR   TIGRFAMs; TIGR00128; fabD; 3.
DR   PROSITE; PS50075; ACP_DOMAIN; 3.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; UNKNOWN_3.
DR   PROSITE; PS00343; GRAM_POS_ANCHORING; UNKNOWN_3.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; UNKNOWN_3.
KW   Phosphopantetheine.
SQ   SEQUENCE   5644 AA;  585775 MW;  CF6A35E3BA0D2658 CRC64;

  Query Match             57.1%;  Score 245;  DB 2;  Length 5644;
  Best Local Similarity   59.8%;  Pred. No. 6e-18;
  Matches   49;  Conservative    7;  Mismatches   26;  Indels    0;  Gaps    0;

Qy          1 LAGQTHEQQHTTLLALVRSHIATVLGHTTPDTIPPDRAFRDLGFDSLTAVELRNRLSRTT 60
              |||   ||:| ||: |||   | ||||:  | :| || |  |||||| ||||| |||  |
Db       5474 LAGLGAEQRHETLVGLVRQETAGVLGHSGADAVPADRDFSRLGFDSLMAVELRTRLSAAT 5533

Qy         61 GLRLPTTLAFDHPNPTTLTHHL 82
              |:|||:|| |||| |   :  ||
Db       5534 GVRLPSTLVFDHPTPAAVARHL 5555


RESULT 13
Q9S0R7
ID   Q9S0R7      PRELIMINARY;      PRT;  6239 AA.
AC   Q9S0R7;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Type I polyketide synthase AVES 2.
GN   AVEA2.
OS   Streptomyces avermitilis.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=33903;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99380548; PubMed=10449723;
RA   Ikeda H., Nonomiya T., Usami M., Ohta T., Omura S.;
RT   "Organization of the biosynthetic gene cluster for the polyketide
RT   anthelmintic macrolide avermectin in Streptomyces avermitilis.";
RL   Proc. Natl. Acad. Sci. U.S.A. 96:9509-9514(1999).
DR   EMBL; AB032367; BAA84475.1; -.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR003880; Ppantne_attach.
DR   InterPro; IPR000169; SHprot_acsite.
DR   InterPro; IPR000834; Zn_carbOpept.
DR   Pfam; PF00698; Acyl_transf; 4.
DR   Pfam; PF00109; ketoacyl-synt; 4.
DR   Pfam; PF02801; ketoacyl-synt_C; 4.
DR   Pfam; PF00550; pp-binding; 4.
DR   PROSITE; PS50075; ACP_DOMAIN; 4.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 4.
DR   PROSITE; PS00133; CARBOXYPEPT_ZN_2; UNKNOWN_4.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; 3.
DR   PROSITE; PS00639; THIOL_PROTEASE_HIS; UNKNOWN_1.
KW   Phosphopantetheine; Transferase.
SQ   SEQUENCE   6239 AA;  666283 MW;  A706280940B502FA CRC64;

  Query Match             57.1%;  Score 245;  DB 2;  Length 6239;
  Best Local Similarity   62.2%;  Pred. No. 6.8e-18;
  Matches   51;  Conservative    7;  Mismatches   24;  Indels    0;  Gaps    0;

Qy          1 LAGQTHEQQHTTLLALVRSHIATVLGHTTPDTIPPDRAFRDLGFDSLTAVELRNRLSRTT 60
              || |:  ::   || |||||:| ||||:  | |   ||||:||||||||||||||:  |
Db       6032 LARQSAGERRQALLRLVRSHVAAVLGHSGADGIDASRAFRELGFDSLTAVELRNRLTAAT 6091

Qy         61 GLRLPTTLAFDHPNPTTLTHHL 82
              ||||  ||||| | |  |  ||
Db       6092 GLRLRATLAFDFPTPAALAEHL 6113


RESULT 14
Q93HJ2
ID   Q93HJ2      PRELIMINARY;      PRT;  4685 AA.
AC   Q93HJ2;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Modular polyketide synthase.
GN   OLMA6.
OS   Streptomyces avermitilis.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=33903;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21477403; PubMed=11572948;
RA   Omura S., Ikeda H., Ishikawa J., Hanamoto A., Takahashi C.,
RA   Shinose M., Takahashi Y., Horikawa H., Nakazawa H., Osonoe T.,
RA   Kikuchi H., Shiba T., Sakaki Y., Hattori M.;
RT   "Genome sequence of an industrial microorganism Streptomyces
RT   avermitilis: Deducing the ability of producing secondary
RT   metabolites.";
RL   Proc. Natl. Acad. Sci. U.S.A. 98:12215-12220(2001).
DR   EMBL; AB070940; BAB69195.1; -.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR002198; ADH_short.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR003880; Ppantne_attach.
DR   Pfam; PF00698; Acyl_transf; 3.
DR   Pfam; PF00106; adh_short; 3.
DR   Pfam; PF00109; ketoacyl-synt; 3.
DR   Pfam; PF02801; ketoacyl-synt_C; 3.
DR   Pfam; PF00550; pp-binding; 3.
DR   PROSITE; PS50075; ACP_DOMAIN; 3.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; UNKNOWN_3.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; UNKNOWN_3.
KW   Phosphopantetheine.
SQ   SEQUENCE   4685 AA;  489355 MW;  2FA34215A39A0DB8 CRC64;

  Query Match             56.9%;  Score 244;  DB 2;  Length 4685;
  Best Local Similarity   58.5%;  Pred. No. 6.3e-18;
  Matches   48;  Conservative    9;  Mismatches   25;  Indels    0;  Gaps    0;
```

```
Qy          1 LAGQTHEQQHTTLLALVRSHIATVLGHTTPDTIPPDRAFRDLGFDSLTAVELRNRLSRTT 60
              ||| |  ::  ||| |||   | || | :   | | | |::|||||||::||||||: |
Db       4518 LAGLTDPERQATLLGLVRGEAAQVLAHGSTAEITPTRPFKELGFDSLTAMDLRNRLSKAT 4577

Qy         61 GLRLPTTLAFDHPNPTTLTHHL 82
              ||||| || ||||||  :  ||
Db       4578 GLRLPATLVFDHPNPQRIAEHL 4599


RESULT 15
Q9F7T9
ID   Q9F7T9      PRELIMINARY;      PRT;  3626 AA.
AC   Q9F7T9;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Avermectin polyketide synthase (Fragment).
OS   Streptomyces avermitilis.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=33903;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC31271;
RA   Hong Y.-S., Lee J.J.;
RT   "Targeted Gene Disruption of the avermectin O-methyltransferase gene
RT   and polyketide synthase gene from Streptomyces avermitilis.";
RL   Submitted (JUN-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF275943; AAG09812.1; -.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR003880; Ppantne_attach.
DR   Pfam; PF00698; Acyl_transf; 3.
DR   Pfam; PF00109; ketoacyl-synt; 2.
DR   Pfam; PF02801; ketoacyl-synt_C; 2.
DR   Pfam; PF00550; pp-binding; 2.
DR   PROSITE; PS50075; ACP_DOMAIN; 2.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 2.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; UNKNOWN_1.
FT   NON_TER    3626    3626
SQ   SEQUENCE   3626 AA;  380557 MW;  6272F5F088C1A8D0 CRC64;

  Query Match             56.6%;  Score 243;  DB 2;  Length 3626;
  Best Local Similarity   59.8%;  Pred. No. 6e-18;
  Matches   49;  Conservative    5;  Mismatches   28;  Indels    0;  Gaps    0;

Qy          1 LAGQTHEQQHTTLLALVRSHIATVLGHTTPDTIPPDRAFRDLGFDSLTAVELRNRLSRTT 60
              || ||    |   || ||| |:| || |. |  | ||::|| |||||||||| || | : |
Db       1946 LAHQTSADQRAALLELVRDHVAAVLRHADPKAIAPDQSFRALGFDSLTAVEFRNLLIKAT 2005

Qy         61 GLRLPTTLAFDHPNPTTLTHHL 82
              ||||| :| |||| |  |  ||
Db       2006 GLRLPVSLVFDHPTPAKLAVHL 2027


Search completed: June 17, 2003, 13:07:58
Job time : 3.78986 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 17, 2003, 12:55:32 ; Search time 1.83105 Seconds
(without alignments)
4305.183 Million cell updates/sec

Title: US-09-914-286-3_COPY_3805_3886
Perfect score: 429
Sequence: 1 LAGQTHEQQHTTLLALVRSH.....RLPTTLAFDHPNPTTLTHHL 82

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_73:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	253	59.0	3739	2	T17410	polyketide synthas
2	249	58.0	4151	2	G70944	probable polyketid
3	246	57.3	1763	2	T17465	rifamycin polyketi
4	237	55.2	1728	2	T17466	rifamycin polyketi
5	234	54.5	1562	2	T17411	polyketide synthas
6	234	54.5	2100	2	T03223	probable polyketid
7	233	54.3	5069	2	T17464	rifamycin polyketi
8	229	53.4	3519	2	S43048	polyketide synthas
9	226	52.7	1762	2	T03222	probable polyketid
10	224	52.2	3413	2	T17467	rifamycin polyketi
11	223	52.0	3491	2	T43231	probable 6-deoxyer
12	218	50.8	10223	2	T30225	polyketide synthas
13	212	49.4	8563	2	T30226	polyketide synthas
14	211	49.2	4735	2	T17463	rifamycin polyketi
15	210	49.0	3573	2	S23070	erythronolide synt
16	209.5	48.8	502	2	A70985	probable polyketid
17	209	48.7	1937	2	T03224	probable polyketid
18	209	48.7	4613	2	T17409	polyketide synthas
19	207	48.3	3172	2	S22012	erythronolide synt
20	207	48.3	3178	2	S13595	6-deoxyerythronoli
21	207	48.3	6260	2	T30228	polyketide synthas
22	203	47.3	6420	2	T30283	polyketide synthas
23	202	47.1	7576	2	T17428	FK506 polyketide s
24	201	46.9	1198	2	T28678	polyketide synthas
25	201	46.9	2103	2	G86925	probable polyketid
26	199	46.4	1616	2	G70668	polyketide synthas
27	197	45.9	2723	2	T03221	probable polyketid
28	196	45.7	2126	2	H70621	probable polyketid
29	189	44.1	1346	2	T17412	polyketide synthas
30	178	41.5	1402	2	D70634	probable polyketid
31	169	39.4	2124	2	T28658	polyketide synthas
32	161.5	37.6	2478	2	AH2140	polyketide synthas
33	155.5	36.2	1017	2	B70985	probable polyketid
34	131	30.5	707	2	F86925	probable acyl-CoA
35	121	28.2	705	2	A70669	probable acid-CoA
36	121	28.2	1540	2	H87203	polyketide synthas
37	116.5	27.2	8243	2	T31307	type I fatty acid
38	114	26.6	1538	2	E70874	probable ppsB prot
39	105	24.5	2518	2	AI2140	polyketide synthas
40	103	24.0	1570	2	AC2012	hypothetical prote
41	102	23.8	2201	2	S73014	polyketide synthas
42	99.5	23.2	2458	2	T17420	probable polyketid
43	99	23.1	130	2	S73018	hypothetical prote
44	96.5	22.5	1582	2	E70876	probable polyketid
45	95.5	22.3	4930	2	E69679	polyketide synthet

ALIGNMENTS

RESULT 1
T17410
polyketide synthase type I - Streptomyces venezuelae
C;Species: Streptomyces venezuelae
C;Date: 02-Sep-2000 #sequence_revision 02-Sep-2000 #text_change 17-Nov-2000
C;Accession: T17410
R;Xue, Y.; Zhao, L.; Liu, H.W.; Sherman, D.H.
Proc. Natl. Acad. Sci. U.S.A. 95, 12111-12116, 1998
A;Title: A gene cluster for macrolide antibiotic biosynthesis in streptomyces venezuela
A;Reference number: Z18773; MUID:98445333; PMID:9770448
A;Accession: T17410
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-3739 <XUE>
A;Cross-references: EMBL:AF079138; NID:g3808326; PID:g3800835; PIDN:AAC69330.1
C;Genetics:
A;Gene: pikAII
C;Superfamily: acyl carrier protein homology
C;Keywords: antibiotic biosynthesis; carrier protein
F;1445-1516/Domain: acyl carrier protein homology <ACP1>
F;3570-3641/Domain: acyl carrier protein homology <ACP2>

Query Match 59.0%; Score 253; DB 2; Length 3739;
Best Local Similarity 62.2%; Pred. No. 3.4e-20;
Matches 51; Conservative 7; Mismatches 24; Indels 0; Gaps 0;

```
Qy          1 LAGQTHEQQHTTLLALVRSHIATVLGHTTPDTIPPDRAFRDLGFDSLTAVELRNRLSRTT 60
              ||  | : :   |  |||:|:|||||| ||  :  :||||| ||||||||||||||:  |
Db       3560 LAAMTPDDRVAHLRDLVRTHVATVLGHGTPSRVDLERAFRDTGFDSLTAVELRNRLNAAT 3619

Qy         61 GLRLPTTLAFDHPNPTTLTHHL 82
              ||||| || |||| |  |  ||
Db       3620 GLRLPATLVFDHPTPGELAGHL 3641
```

RESULT 2
G70944
probable polyketide synthase Rv2048c - Mycobacterium tuberculosis (strain H37RV)
C;Species: Mycobacterium tuberculosis
C;Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
C;Accession: G70944
R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon,
; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A;Authors: Sqares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A;Reference number: A70500; MUID:98295987; PMID:9634230
A;Accession: G70944
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-4151 <COL>

A;Cross-references: GB:AL021899; GB:AL123456; NID:g3242282; PIDN:CAA17262.1; PID:g289678
A;Experimental source: strain H37Rv
C;Genetics:
A;Gene: pks12
C;Superfamily: Mycobacterium tuberculosis probable polyketide synthase Rv2048c; 3-oxoacy
homology; [acyl-carrier-protein] S-malonyltransferase homology
C;Keywords: carrier protein
F;56-455/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS>
F;559-841/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT>
F;1680-1861/Domain: short-chain alcohol dehydrogenase homology <SAD1>
F;1963-2038/Domain: acyl carrier protein homology <ACP1>
F;2078-2478/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS1>
F;2582-2860/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT1>
F;3710-3891/Domain: short-chain alcohol dehydrogenase homology <SAD2>
F;3995-4066/Domain: acyl carrier protein homology <ACP2>

Query Match 58.0%; Score 249; DB 2; Length 4151;
Best Local Similarity 59.8%; Pred. No. 1.1e-19;
Matches 49; Conservative 9; Mismatches 24; Indels 0; Gaps 0;

Qy 1 LAGQTHEQQHTTLLALVRSHIATVLGHTTPDTIPPDRAFRDLGFDSLTAVELRNRLSRTT 60
Db 1953 LHGLPEAEQHAVLLGLVRLHIATVLGNITPEAIDPDKAFQDLGFDSLTAVEMRNRLKSAT 2012

Qy 61 GLRLPTTLAFDHPNPTTLTHHL 82
Db 2013 GLSLSPTLIFDYPTPNRLASYI 2034

RESULT 3
T17465
rifamycin polyketide synthase module 7 - Amycolatopsis mediterranei
C;Species: Amycolatopsis mediterranei
C;Date: 02-Sep-2000 #sequence_revision 02-Sep-2000 #text_change 03-Nov-2000
C;Accession: T17465
R;Schupp, T.
submitted to the EMBL Data Library, December 1997
A;Reference number: Z18802
A;Accession: T17465
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1763 <SCH>
A;Cross-references: EMBL:AJ223012; NID:e1227119; PID:e1227122; PIDN:CAA11037.1
A;Experimental source: strain LBG A3136
C;Superfamily: Streptomyces hygroscopicus probable polyketide synthase module 4; 3-oxoac
homology; [acyl-carrier-protein] S-malonyltransferase homology
C;Keywords: carrier protein
F;1615-1686/Domain: acyl carrier protein homology <ACP>

Query Match 57.3%; Score 246; DB 2; Length 1763;
Best Local Similarity 59.8%; Pred. No. 8.7e-20;
Matches 49; Conservative 6; Mismatches 27; Indels 0; Gaps 0;

Qy 1 LAGQTHEQQHTTLLALVRSHIATVLGHTTPDTIPPDRAFRDLGFDSLTAVELRNRLSRTT 60
Db 1605 LAGLTVAEQEALLLDLVRGQVAVVLGHADSSGVRADAAFKDAGFDSLTSVELRNRLRETT 1664

Qy 61 GLRLPTTLAFDHPNPTTLTHHL 82
Db 1665 GLKLPATLVFDHPNPLALARHL 1686

RESULT 4
T17466
rifamycin polyketide synthase module 8 - Amycolatopsis mediterranei
C;Species: Amycolatopsis mediterranei
C;Date: 02-Sep-2000 #sequence_revision 02-Sep-2000 #text_change 03-Nov-2000
C;Accession: T17466
R;Schupp, T.
submitted to the EMBL Data Library, December 1997
A;Reference number: Z18802
A;Accession: T17466
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1728 <SCH>
A;Cross-references: EMBL:AJ223012; NID:e1227119; PID:e1227123; PIDN:CAA11038.1
A;Experimental source: strain LBG A3136
C;Superfamily: Streptomyces hygroscopicus probable polyketide synthase module 4; 3-oxoa
homology; [acyl-carrier-protein] S-malonyltransferase homology
C;Keywords: carrier protein
F;1573-1644/Domain: acyl carrier protein homology <ACP>

Query Match 55.2%; Score 237; DB 2; Length 1728;
Best Local Similarity 56.1%; Pred. No. 9.3e-19;
Matches 46; Conservative 10; Mismatches 26; Indels 0; Gaps 0;

Qy 1 LAGQTHEQQHTTLLALVRSHIATVLGHTTPDTIPPDRAFRDLGFDSLTAVELRNRLSRTT 60
Db 1563 LAGLGAPEQEALLVDLVRGQVAAVLGHAGPDAVRADTAFKDAGFDSLTSVDLRNRLREST 1622

Qy 61 GLRLPTTLAFDHPNPTTLTHHL 82
Db 1623 GLKLPATLAFDYPTPLVLARHL 1644

RESULT 5
T17411
polyketide synthase III - Streptomyces venezuelae
C;Species: Streptomyces venezuelae
C;Date: 02-Sep-2000 #sequence_revision 02-Sep-2000 #text_change 03-Nov-2000
C;Accession: T17411
R;Xue, Y.; Zhao, L.; Liu, H.W.; Sherman, D.H.
Proc. Natl. Acad. Sci. U.S.A. 95, 12111-12116, 1998
A;Title: A gene cluster for macrolide antibiotic biosynthesis in streptomyces venezuela
A;Reference number: Z18773; MUID:98445333; PMID:9770448
A;Accession: T17411
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1562 <XUE>
A;Cross-references: EMBL:AF079138; NID:g3808326; PID:g3800836; PIDN:AAC69331.1
C;Genetics:
A;Gene: pikAIII
C;Superfamily: Streptomyces hygroscopicus probable polyketide synthase module 4; 3-oxoa
homology; [acyl-carrier-protein] S-malonyltransferase homology
C;Keywords: antibiotic biosynthesis; carrier protein
F;1403-1474/Domain: acyl carrier protein homology <ACP>

Query Match 54.5%; Score 234; DB 2; Length 1562;
Best Local Similarity 58.7%; Pred. No. 1.8e-18;
Matches 44; Conservative 9; Mismatches 22; Indels 0; Gaps 0;

Qy 8 QQHTTLLALVRSHIATVLGHTTPDTIPPDRAFRDLGFDSLTAVELRNRLSRTTGLRLPTT 67
Db 1400 ERRPALLTLVRTHAAAVLGHSSPDRVAPGRAFTELGFDSLTAVQLRNQLSTVVGNRLPAT 1459

Qy 68 LAFDHPNPTTLTHHL 82
Db 1460 TVFDHPTPAALAAHL 1474

RESULT 6
T03223
probable polyketide synthase module 3 - Streptomyces hygroscopicus
C;Species: Streptomyces hygroscopicus
C;Date: 24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change 26-May-2000
C;Accession: T03223
R;Ruan, X.; Stassi, D.; Lax, S.; Katz, L.
Gene 203, 1-9, 1997
A;Title: A second type-I PKS gene cluster isolated from Streptomyces hygroscopicus ATCC
A;Reference number: Z14848; MUID:98085969; PMID:9426000
A;Accession: T03223
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-2100 <RUA>
A;Cross-references: EMBL:AF007101; NID:g2624946; PIDN:AAC38063.1; PID:g2624950
A;Experimental source: ATCC 29253

```
C;Superfamily: mycocerosic acid synthase; 3-oxoacyl-[acyl-carrier-protein] synthase I ho
nase homology; [acyl-carrier-protein] S-malonyltransferase homology
C;Keywords: carrier protein
F;55-456/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS>
F;566-851/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT>
F;1952-2027/Domain: acyl carrier protein homology <ACP>

  Query Match             54.5%;  Score 234;  DB 2;  Length 2100;
  Best Local Similarity   54.9%;  Pred. No. 2.6e-18;
  Matches   45;  Conservative   12;  Mismatches   25;  Indels    0;  Gaps    0;

Qy          1 LAGQTHEQQHTTLLALVRSHIATVLGHTTPDTIPPDRAFRDLGFDSLTAVELRNRLSRTT 60
              :|    |::   :| |||:  ||||||:  : : |:||:| |||||||||| |||::  |
Db       1942 VAAVAPEERRAVVLELVRAQAATVLGHSGAEHVSPTQAFKDSGFDSLTAVEFRNRVTTAT 2001

Qy         61 GLRLPTTLAFDHPNPTTLTHHL 82
              ||||| || |||| | ||  ||
Db       2002 GLRLPATLIFDHPTPNTLADHL 2023


RESULT 7
T17464
rifamycin polyketide synthase modules 4-6 - Amycolatopsis mediterranei
C;Species: Amycolatopsis mediterranei
C;Date: 02-Sep-2000 #sequence_revision 02-Sep-2000 #text_change 17-Nov-2000
C;Accession: T17464
R;Schupp, T.
submitted to the EMBL Data Library, December 1997
A;Reference number: Z18802
A;Accession: T17464
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-5069 <SCH>
A;Cross-references: EMBL:AJ223012; NID:e1227119; PID:e1227121; PIDN:CAA11036.1
A;Experimental source: strain LBG A3136
C;Superfamily: acyl carrier protein homology
C;Keywords: carrier protein
F;1631-1702/Domain: acyl carrier protein homology <ACP1>
F;3238-3309/Domain: acyl carrier protein homology <ACP2>
F;4939-5010/Domain: acyl carrier protein homology <ACP3>

  Query Match             54.3%;  Score 233;  DB 2;  Length 5069;
  Best Local Similarity   57.3%;  Pred. No. 1e-17;
  Matches   47;  Conservative    7;  Mismatches   28;  Indels    0;  Gaps    0;

Qy          1 LAGQTHEQQHTTLLALVRSHIATVLGHTTPDTIPPDRAFRDLGFDSLTAVELRNRLSRTT 60
              |||     | | || |||: :| ||||     :  | ||:| ||||||||||||:   |
Db       3228 LAGLAPADQLTLLLDLVRAQVAAVLGHADASAVRVDTAFKDAGFDSLTAVELRNRMRTAT 3287

Qy         61 GLRLPTTLAFDHPNPTTLTHHL 82
              ||:|| || ||:|||  |  ||
Db       3288 GLKLPATLVFDYPNPQALARHL 3309


RESULT 8
S43048
polyketide synthase type I - Streptomyces antibioticus
N;Contains: acyl carrier protein; acyltransferase; ketoacylsynthase; ketoreductase; thio
C;Species: Streptomyces antibioticus
C;Date: 13-Jan-1995 #sequence_revision 06-Dec-1996 #text_change 26-May-2000
C;Accession: S43048; S41729
R;Swan, D.G.; Rodriguez, A.M.; Vilches, C.; Mendez, C.; Salas, J.A.
submitted to the EMBL Data Library, February 1993
A;Reference number: S43048
A;Accession: S43048
A;Molecule type: DNA
A;Residues: 1-3519 <SWA>
A;Cross-references: EMBL:L09654; NID:g153407; PIDN:AAA19695.1; PID:g153408
R;Swan, D.G.; Rodriguez, A.M.; Vilches, C.; Mendez, C.; Salas, J.A.
Mol. Gen. Genet. 242, 358-362, 1994
A;Title: Characterisation of a Streptomyces antibioticus gene encoding a type I polyketi
A;Reference number: S41729; MUID:94150470; PMID:8107683
A;Accession: S41729
A;Status: nucleic acid sequence not shown
A;Molecule type: DNA
A;Residues: 1683-3238,3273-3303,'K',3305-3407,'T',3409-3462,'Y',3464-3516,'E',3518-3519
A;Cross-references: EMBL:L09654
C;Genetics:
A;Start codon: GTG
C;Superfamily: [acyl-carrier-protein] S-malonyltransferase homology; 3-oxoacyl-[acyl-ca
ogy; short-chain alcohol dehydrogenase homology
C;Keywords: antibiotic biosynthesis; carrier protein
F;59-462/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS1>
F;570-851/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT1>
F;1202-1381/Domain: short-chain alcohol dehydrogenase homology <SAD1>
F;1489-1560/Domain: acyl carrier protein homology <ACP1>
F;1708-2111/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS2>
F;2221-2502/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT2>
F;2858-3037/Domain: short-chain alcohol dehydrogenase homology <SAD2>
F;3143-3214/Domain: acyl carrier protein homology <ACP2>
F;3305-3500/Domain: oleoyl-[acyl-carrier-protein] hydrolase homology <ACPH>

  Query Match             53.4%;  Score 229;  DB 2;  Length 3519;
  Best Local Similarity   53.7%;  Pred. No. 1.9e-17;
  Matches   44;  Conservative   14;  Mismatches   24;  Indels    0;  Gaps    0;

Qy          1 LAGQTHEQQHTTLLALVRSHIATVLGHTTPDTIPPDRAFRDLGFDSLTAVELRNRLSRTT 60
              |:| :  :|    |: |||:  | |||| :    :| :|||::|||||||||||||| |:  |
Db       1479 LSGLSRVRQEEELVELVRAQAAVVLGHGSAQDVPAERAFKELGFDSLTAVELRNGLAAAT 1538

Qy         61 GLRLPTTLAFDHPNPTTLTHHL 82
              |:||| |:|||||| | :   |
Db       1539 GIRLPATMAFDHPNATAIARFL 1560


RESULT 9
T03222
probable polyketide synthase module 2 - Streptomyces hygroscopicus
C;Species: Streptomyces hygroscopicus
C;Date: 24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change 03-Nov-2000
C;Accession: T03222
R;Ruan, X.; Stassi, D.; Lax, S.; Katz, L.
Gene 203, 1-9, 1997
A;Title: A second type-I PKS gene cluster isolated from Streptomyces hygroscopicus ATCC
A;Reference number: Z14848; MUID:98085969; PMID:9426000
A;Accession: T03222
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1762 <RUA>
A;Cross-references: EMBL:AF007101; NID:g2624946; PIDN:AAC38062.1; PID:g2624949
A;Experimental source: ATCC 29253
C;Superfamily: Streptomyces hygroscopicus probable polyketide synthase module 4; 3-oxoa
 homology; [acyl-carrier-protein] S-malonyltransferase homology
C;Keywords: carrier protein
F;54-454/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS>
F;550-822/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT>
F;1598-1669/Domain: acyl carrier protein homology <ACP>

  Query Match             52.7%;  Score 226;  DB 2;  Length 1762;
  Best Local Similarity   54.9%;  Pred. No. 1.8e-17;
  Matches   45;  Conservative    9;  Mismatches   28;  Indels    0;  Gaps    0;

Qy          1 LAGQTHEQQHTTLLALVRSHIATVLGHTTPDTIPPDRAFRDLGFDSLTAVELRNRLSRTT 60
              :|    |::   :| |||   | ||||:  | | | :||:| |||||||||| |||::  |
Db       1588 IAALMPEERDRAVLDLVRQQSARVLGHSDADRIDPGQAFKDTGFDSLTAVEFRNRVTTAT 1647

Qy         61 GLRLPTTLAFDHPNPTTLTHHL 82
              ||||| || |||| |  |  ||
Db       1648 GLRLPATLIFDHPTPNALADHL 1669


RESULT 10
T17467
rifamycin polyketide synthase modules 9-10 - Amycolatopsis mediterranei
C;Species: Amycolatopsis mediterranei
C;Date: 02-Sep-2000 #sequence_revision 02-Sep-2000 #text_change 17-Nov-2000
```

```
C;Accession: T17467
R;Schupp, T.
submitted to the EMBL Data Library, December 1997
A;Reference number: Z18802
A;Accession: T17467
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-3413 <SCH>
A;Cross-references: EMBL:AJ223012; NID:e1227119; PID:e1227124; PIDN:CAA11039.1
A;Experimental source: strain LBG A3136
C;Superfamily: acyl carrier protein homology
C;Keywords: carrier protein
F;1608-1679/Domain: acyl carrier protein homology <ACP1>
F;3334-3405/Domain: acyl carrier protein homology <ACP2>

  Query Match             52.2%;  Score 224;  DB 2;  Length 3413;
  Best Local Similarity   52.4%;  Pred. No. 6.8e-17;
  Matches   43;  Conservative   12;  Mismatches   27;  Indels    0;  Gaps    0;

Qy          1 LAGQTHEQQHTTLLALVRSHIATVLGHTTPDTIPPDRAFRDLGFDSLTAVELRNRLSRTT 60
              |||    :|   || :||: :| ||||  |: :  | ||:| ||||||:|||||||   :
Db       1598 LAGLAPAEQEALLLDVVRTQVALVLGHAGPEAVRADTAFKDTGFDSLTSVELRNRLREAS 1657

Qy         61 GLRLPTTLAFDHPNPTTLTHHL 82
              ||:|| || ||:| |  |  :|
Db       1658 GLKLPATLVFDYPTPVALARYL 1679


RESULT 11
T43231
probable 6-deoxyerythronolide B synthase, module 1/2 - Saccharopolyspora erythraea
C;Species: Saccharopolyspora erythraea
C;Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 01-Dec-2000
C;Accession: T43231
R;Donadio, S.; Staver, M.J.; McAlpine, J.B.; Swanson, S.J.; Katz, L.
Science 252, 675-679, 1991
A;Title: Modular organization of genes required for complex polyketide biosynthesis.
A;Reference number: Z22354; MUID:91220065; PMID:2024119
A;Accession: T43231
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-3491 <DON>
A;Cross-references: EMBL:M63676; NID:g152691; PID:g152692; PIDN:AAA26493.1
C;Genetics:
A;Gene: eryA
C;Superfamily: 3-oxoacyl-[acyl-carrier-protein] synthase I homology; acyl carrier protei
C;Keywords: carrier protein; phosphopantetheine; phosphoprotein
F;57-336/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT1>
F;416-483/Domain: acyl carrier protein homology <ACP1>
F;526-926/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS1>
F;1031-1317/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT2>
F;1886-1957/Domain: acyl carrier protein homology <ACP2>
F;2001-2400/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS2>
F;2508-2791/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT3>
F;3332-3403/Domain: acyl carrier protein homology <ACP3>
F;1921/Binding site: phosphopantetheine (Ser) (covalent) #status predicted

  Query Match             52.0%;  Score 223;  DB 2;  Length 3491;
  Best Local Similarity   53.7%;  Pred. No. 9.1e-17;
  Matches   44;  Conservative   11;  Mismatches   27;  Indels    0;  Gaps    0;

Qy          1 LAGQTHEQQHTTLLALVRSHIATVLGHTTPDTIPPDRAFRDLGFDSLTAVELRNRLSRTT 60
              ||     ::   |  ||||| | |||| : : :| |:|| :|| |||:|:||||||   |
Db       1876 LASLPAPEREEALFELVRSHAAAVLGHASAERVPADQAFAELGVDSLSALELRNRLGAAT 1935

Qy         61 GLRLPTTLAFDHPNPTTLTHHL 82
              |:|||||  ||||:  ||  ||
Db       1936 GVRLPTTTVFDHPDVRTLAAHL 1957


RESULT 12
T30225
polyketide synthase - Streptomyces hygroscopicus
C;Species: Streptomyces hygroscopicus
C;Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 17-Nov-2000
C;Accession: T30225
R;Aparicio, J.F.; Molnar, I.; Schwecke, T.; Koenig, A.; Haydock, S.F.; Khaw, L.E.; Staun
Gene 169, 9-16, 1996
A;Title: Organization of the biosynthetic gene cluster for rapamycin in Streptomyces hyg
A;Reference number: Z20782; MUID:96186896; PMID:8635756
A;Accession: T30225
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-10223 <APA>
A;Cross-references: EMBL:X86780; NID:g987088; PID:g987099; PIDN:CAA60459.1
C;Genetics:
A;Gene: rapB
C;Superfamily: 3-oxoacyl-[acyl-carrier-protein] synthase I homology; acyl carrier protei
C;Keywords: carrier protein
F;54-449/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS1>
F;1533-1604/Domain: acyl carrier protein homology <ACP1>
F;1647-2042/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS2>
F;2135-2409/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT1>
F;3172-3243/Domain: acyl carrier protein homology <ACP2>
F;3287-3681/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS3>
F;3778-4052/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT2>
F;5160-5231/Domain: acyl carrier protein homology <ACP3>
F;5275-5667/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS4>
F;6776-6847/Domain: acyl carrier protein homology <ACP4>
F;6891-7285/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS5>
F;8411-8482/Domain: acyl carrier protein homology <ACP5>
F;8526-8921/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS6>
F;9012-9285/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT3>
F;10069-10140/Domain: acyl carrier protein homology <ACP>

  Query Match             50.8%;  Score 218;  DB 2;  Length 10223;
  Best Local Similarity   57.3%;  Pred. No. 1.3e-15;
  Matches   47;  Conservative    6;  Mismatches   29;  Indels    0;  Gaps    0;

Qy          1 LAGQTHEQQHTTLLALVRSHIATVLGHTTPDTIPPDRAFRDLGFDSLTAVELRNRLSRTT 60
              ||    |::   || :||   ||||||    |||   |||||| |||||||||| |:: |
Db       1523 LAALAPEERAKALLKVVRDTAATVLGHADARTIPVTGAFRDLGIDSLTAVELRNGLAKVT 1582

Qy         61 GLRLPTTLAFDHPNPTTLTHHL 82
              ||||| || ||:| |  |    |
Db       1583 GLRLPATLVFDYPTPAVLAARL 1604


RESULT 13
T30226
polyketide synthase - Streptomyces hygroscopicus
C;Species: Streptomyces hygroscopicus
C;Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 01-Dec-2000
C;Accession: T30226
R;Aparicio, J.F.; Molnar, I.; Schwecke, T.; Koenig, A.; Haydock, S.F.; Khaw, L.E.; Staun
Gene 169, 9-16, 1996
A;Title: Organization of the biosynthetic gene cluster for rapamycin in Streptomyces hyg
A;Reference number: Z20782; MUID:96186896; PMID:8635756
A;Accession: T30226
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-8563 <APA>
A;Cross-references: EMBL:X86780; NID:g987088; PID:g987100; PIDN:CAA60460.1
C;Genetics:
A;Gene: rapA
C;Superfamily: 3-oxoacyl-[acyl-carrier-protein] synthase I homology; acetate-CoA ligase
C;Keywords: carrier protein; phosphopantetheine; phosphoprotein
F;54-503/Domain: acetate-CoA ligase homology <ACL>
F;1329-1724/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS1>
F;1817-2091/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT1>
F;3199-3270/Domain: acyl carrier protein homology <ACP1>
F;3314-3706/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS2>
F;4787-4858/Domain: acyl carrier protein homology <ACP2>
F;4902-5293/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS3>
F;5386-5659/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT2>
F;6760-6831/Domain: acyl carrier protein homology <ACP3>
```

```
F;6875-7269/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS4>
F;7362-7636/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT3>
F;8412-8483/Domain: acyl carrier protein homology <ACP4>
F;4822/Binding site: phosphopantetheine (Ser) (covalent) #status predicted

  Query Match             49.4%;  Score 212;  DB 2;  Length 8563;
  Best Local Similarity   54.9%;  Pred. No. 5e-15;
  Matches   45;  Conservative    9;  Mismatches   28;  Indels    0;  Gaps    0;

Qy          1 LAGQTHEQQHTTLLALVRSHIATVLGHTTPDTIPPDRAFRDLGFDSLTAVELRNRLSRTT 60
              ||    |::   || :|    ||||||   |:||   ||:||| |||||||||| |:: |
Db       8402 LAALAPEERAKALLRVVCDSAATVLGHADIDSIPVTAAFKDLGVDSLTAVDLRNSLAKAT 8461

Qy         61 GLRLPTTLAFDHPNPTTLTHHL 82
              ||||| || ||:| || |   |
Db       8462 GLRLPPTLVFDYPTPTALAARL 8483


RESULT 14
T17463
rifamycin polyketide synthase modules 1-3 - Amycolatopsis mediterranei
C;Species: Amycolatopsis mediterranei
C;Date: 02-Sep-2000 #sequence_revision 02-Sep-2000 #text_change 17-Nov-2000
C;Accession: T17463
R;Schupp, T.
submitted to the EMBL Data Library, December 1997
A;Reference number: Z18802
A;Accession: T17463
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-4735 <SCH>
A;Cross-references: EMBL:AJ223012; NID:e1227119; PID:e1227120; PIDN:CAA11035.1
A;Experimental source: strain LBG A3136
C;Superfamily: acyl carrier protein homology; acetate-CoA ligase homology
C;Keywords: carrier protein
F;53-500/Domain: acetate-CoA ligase homology <ACL>
F;543-610/Domain: acyl carrier protein homology <ACP1>
F;2102-2173/Domain: acyl carrier protein homology <ACP2>
F;3079-3150/Domain: acyl carrier protein homology <ACP3>
F;4578-4649/Domain: acyl carrier protein homology <ACP4>

  Query Match             49.2%;  Score 211;  DB 2;  Length 4735;
  Best Local Similarity   52.4%;  Pred. No. 3.2e-15;
  Matches   43;  Conservative    9;  Mismatches   30;  Indels    0;  Gaps    0;

Qy          1 LAGQTHEQQHTTLLALVRSHIATVLGHTTPDTIPPDRAFRDLGFDSLTAVELRNRLSRTT 60
              |||    :|   || ||| | | |||||:  :::   | |:| ||||||||||||||:  |
Db       2092 LAGLAETEQAAALLDLVRRHAAEVLGHSGAESVHSGRTFKDAGFDSLTAVELRNRLAAAT 2151

Qy         61 GLRLPTTLAFDHPNPTTLTHHL 82
              || |   : ||:| |  |  ||
Db       2152 GLTLSPAMIFDYPKPPALADHL 2173


RESULT 15
S23070
erythronolide synthase (EC 2.3.1.94) II - Saccharopolyspora erythraea
N;Alternate names: 6-deoxyerythronolide B synthase II
C;Species: Saccharopolyspora erythraea
C;Date: 07-Apr-1994 #sequence_revision 07-Apr-1994 #text_change 21-Jul-2000
C;Accession: S23070; S22011; S23205
R;Bevitt, D.J.; Cortes, J.; Haydock, S.F.; Leadlay, P.F.
Eur. J. Biochem. 204, 39-49, 1992
A;Title: 6-Deoxyerythronolide-B synthase 2 from Saccharopolyspora erythraea. Cloning of
A;Reference number: S23070; MUID:92155230; PMID:1740151
A;Accession: S23070
A;Molecule type: DNA
A;Residues: 1-3573 <BEV1>
A;Cross-references: EMBL:X62569
A;Experimental source: strain NRRL 2338
R;Bevitt, D.J.
submitted to the EMBL Data Library, September 1991
A;Reference number: S22011
A;Accession: S22011
A;Molecule type: DNA
A;Residues: 1-184,'I',186-301,'S',303-521,523-658,'A',660-993,1001-1212,'H',1214-1392,1
3479,'DH',3480-3572 <BEV2>
A;Cross-references: EMBL:X62569; NID:g46977; PIDN:CAA44448.1; PID:g581651
R;Caffrey, P.; Bevitt, D.J.; Staunton, J.; Leadlay, P.F.
FEBS Lett. 304, 225-228, 1992
A;Title: Identification of DEBS 1, DEBS 2 and DEBS 3, the multienzyme polypeptides of t
A;Reference number: S23103; MUID:92316235; PMID:1618327
A;Accession: S23205
A;Molecule type: protein
A;Residues: 2-12,'XXX' <CAF>
A;Experimental source: strain CA340
C;Genetics:
A;Gene: eryA
A;Start codon: GTG
C;Function:
A;Description: catalyzes the construction of a polyketide chain, which is then cyclised
A;Pathway: erythromycin biosynthesis
C;Superfamily: [acyl-carrier-protein] S-malonyltransferase homology; 3-oxoacyl-[acyl-ca
ort-chain alcohol dehydrogenase homology
C;Keywords: acyltransferase; antibiotic biosynthesis; carrier protein; coenzyme A; mult
F;52-453/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS1>
F;561-843/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT1>
F;1140-1308/Domain: short-chain alcohol dehydrogenase homology <SAD2>
F;1404-1475/Domain: acyl carrier protein homology <ACP1>
F;1519-1919/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS2>
F;2023-2305/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT2>
F;2857-3131/Domain: long-chain alcohol dehydrogenase homology <LADH>
F;3149-3327/Domain: short-chain alcohol dehydrogenase homology <SADH>
F;3420-3493/Domain: acyl carrier protein homology <ACP2>
F;1439/Binding site: phosphopantetheine (Ser) (covalent) #status predicted

  Query Match             49.0%;  Score 210;  DB 2;  Length 3573;
  Best Local Similarity   53.7%;  Pred. No. 2.9e-15;
  Matches   44;  Conservative   10;  Mismatches   26;  Indels    2;  Gaps    1;

Qy          1 LAGQTHEQQHTTLLALVRSHIATVLGHTTPDTIPPDRAFRDLGFDSLTAVELRNRLSRTT 60
              |||::  |   |  ||||| | | |: : | :|  :||:||||||| ||||||||   |
Db       3410 LAGRSESDQVAGLAELVRSHAAAVSGYGSADQLPERKAFKDLGFDSLAAVELRNRLGTAT 3469

Qy         61 GLRLPTTLAFDHPNPTTLTHHL 82
              |:|||:|| |  | |  :  ||
Db       3470 GVRLPSTLVF--PTPLAVAEHL 3489


Search completed: June 17, 2003, 13:12:26
Job time : 2.83105 secs
```

THIS PAGE BLANK (USPTO)